Disseminating Information on Twitter: Evidence from Investment Advisers

Seyed Mohammad Kazempour[a]

[a] *Jones Graduate School of Business*
Rice University, Houston, TX 77005, U.S.A.

Abstract

I show that investment advisers disseminate valuable information about stocks on their Twitter accounts. A one standard deviation increase in sentiment predicts 12 bps abnormal returns over the next week. Advisers' tweets interpret public news, especially analyst revisions and earnings announcements, and also disclose novel information. Advisers offering financial planning services post more informative tweets. Moreover, retail investors trade in the direction of tweets over the following week.

☆I thank my dissertation committee Kerry Back (chair), Bruce Carlin, Richard Boylan, Alan Crane, and Kevin Crotty for their advice. I am grateful to Yessenia Tellez, Mohammad Ghaderi, Ehsan Azarmsa, and seminar participants at Rice for their helpful comments. I also thank Craig Holden for providing his code.

October 6, 2022

1. Introduction

A growing body of papers on information crowdsourcing platforms suggests they can predict stock returns (Chen et al. (2014), Dim (2020), Crawford et al. (2017), Crawford et al. (2018)). For example, Chen et al. (2014) find that the sentiment of Seeking Alpha articles is positively correlated with the abnormal returns of the mentioned stocks over the next three months. The case of Twitter, however, is curious. While some authors have found that Twitter sentiment predicts abnormal returns of stocks with a positive sign (Ballinari and Behrendt (2021), Bartov et al. (2018)), others find the correlation negative (Giannini et al. (2018)). While the value of Twitter is unclear in general, I focus on an important group of finance professionals, registered investment advisers (RIAs), who advised 60.8 million clients on $110 trillion of their assets in 2020, and show that their tweets are informative.

I collect investment advisers' Twitter accounts from their annual form ADV submissions and extract tweets in which they mention stocks. After manually labeling a sample of tweets as positive, neutral, or negative, I train and test nine state-of-the-art neural network models to predict tweet sentiment. For my main analysis, I combine these nine predictions to obtain an ensemble sentiment model. A one standard deviation increase in the sentiment of advisers' tweets predicts 12 bps (6% per annum) of abnormal returns over the next trading week. Moreover, the abnormal returns continue to grow over the next three months, which alleviates the concern that the predictive power of tweets might be due to temporary price pressures or advisers running pump-and-dump schemes.

The result that investment advisors' tweets are informative is somewhat surprising in light of earlier literature that documents advisors' poor portfolio performance. For example, Chalmers and Reuter (2020) show that advisers' clients earn lower after-fee returns and Sharpe ratios than target date funds. Moreover, Foerster et al. (2017) find that adviser-managed portfolios underperform life-cycle funds by 1.5% per year. While advisers' poor performance indicates that their strategies are not valuable in general, my results show that the selection of information they choose to tweet does have value.

My results complement other papers that have found public statements about stocks by finance professionals to be informative. In particular, extensive literature studies the informativeness of analyst recommendations. For example, Loh and Stulz (2010) report that two-day abnormal returns after single-notch upgrades (downgrades) by analysts average to 2.50% (-3.55%). Furthermore, Crane and Crotty (2020) find that 97% of all analysts issue valuable recommendations. Compared to analyst recommendations which are accessible to subscribers, the tweets in my data are available to all investors free of charge.

Tweet informativeness can vary across stocks. I consider two hypotheses. First, the most frequently-tweeted stocks are often popular topics of casual conversation among investors. Therefore, it is reasonable to expect that tweets about the most frequently-tweeted stocks are less informative. Second, consistent with Crane and Crotty (2020) who find that abnormal returns after analyst recommendations are negatively correlated with firm size, I hypothesize that tweets about large stocks are also less informative. To test these hypotheses, I repeat my regressions for three subsamples of stocks: (1) the top 10 most frequently tweeted stocks, (2) S&P500

stocks other than those in the first category, and (3) all stocks not included in the first two categories. Consistent with my hypotheses, tweet sentiment predicts abnormal returns only for the third subsample.

Investment advisers' tweets can interpret recently released public news or contain novel information. To determine the source of advisers' information, I compare the probability of tweeting on news and non-news days. Tweets are more likely following analyst recommendations and earnings announcements than other news categories. I test the predictive power of tweets on news and non-news days separately. Tweet sentiment predicts abnormal returns even on non-news days, suggesting that advisers' tweets sometimes contain novel information. Moreover, advisers' tweets predict returns on news days even after controlling for news sentiment, suggesting they help interpret the prevailing news stories.

Given the heterogeneity in advisers' businesses, it is reasonable to ask which advisers' tweets are more informative. The answer, however, is not clear ex ante. On the one hand, advisers who manage funds are better incentivized to acquire information about stocks. But once the information is acquired, the same incentives discourage them from disclosing their profitable signals. On the other hand, Gurun et al. (2018) show that retail investors withdrew fewer assets from advisers who provide financial planning services following the Madoff scandal, suggesting they trust financial planners more than other advisers. Such advisers may be more likely to disclose profitable signals as part of their trust-building process. In return regressions, I find that financial planners post informative tweets. In contrast, fund managers' tweets do not significantly predict next week's abnormal returns.

Finally, I show that retail investors tend to trade in the direction of tweets. Using the measure of retail order imbalance in Boehmer et al. (2021), I show that a one standard deviation increase in tweet sentiment predicts a 2.6% of a standard deviation increase in retail order imbalance over the following trading week. There are causal and endogenous explanations for this finding. On the one hand, some retail investors may trade on advisers' tweets. On the other hand, one endogenous explanation is that advisers predict the retail trading activity and time their tweets accordingly. Without estimating the relative contributions of these channels, I argue that investment advisers can benefit retail investors by curating their profitable trades.

My paper relates to a branch of literature on crowdsourcing information in financial markets (Chen et al. (2014), Campbell et al. (2019), Crawford et al. (2017), Crawford et al. (2018)). As a prominent example of this literature, Chen et al. (2014) find that the sentiment of articles posted on Seeking Alpha predicts abnormal returns of stocks over the next three months. More specifically, the literature on Twitter has studied an array of topics such as the informativeness of local and nonlocal users' tweets (Giannini et al. (2018)), disagreement among Twitter users (Cookson and Niessner (2020)), information siloing (Cookson et al. (2021)), and the predictive power of tweets before news releases (Bartov et al. (2018), Campbell et al. (2021)). My contribution to this literature is two-fold. First, I show that Twitter can help interpret the news as well as provide novel information. Second, my paper is the first to study information dissemination by a group of finance professionals on Twitter.

I also contribute to a branch of literature on the value of investment advisers. Previous literature has mainly focused on conflicts of interest between advisers and their clients, concluding that adviser-directed investments often underperform passive alternatives (Chalmers and Reuter (2020), Bergstresser et al. (2008), Hackethal et al. (2010)). More recently, Foerster et al. (2017) confirm this finding using data from four Canadian advisers. In a follow-up paper, Linnainmaa et al. (2021) show that advisers' trade as they advise their clients and their portfolios have similar performances. They conclude that advisers' misguided beliefs contribute to the excess trading and underperformance of their clients. My paper contributes to this literature by highlighting advisers' ability to provide profitable signals. Moreover, I provide evidence that advisers are heterogeneous in the amount of information they share.

2. Data

2.1. Twitter

Investment advisers must submit annual amendments to their form ADVs within 90 days of the end of their fiscal year. Effective October 1, 2017, and as part of a revision of form ADV, the SEC requires registered investment advisers to disclose all social media accounts for which they control the content. Here is an excerpt from item 1.I of form ADV:

> "Do you have one or more websites or accounts on publicly available social media platforms (including, but not limited to, Twitter, Facebook and LinkedIn)?

> *If "yes," list all firm website addresses and the address for each of the firm's accounts on publicly available social media platforms on Section 1.I. of Schedule D.*"

After downloading form ADV data from the SEC website through the end of 2020, I extracted 3615 public Twitter accounts belonging to registered investment advisers. The data indicate that Twitter is very popular among investment advisers. As a comparison, I found only seven accounts in Stocktwits and 12 in Seeking Alpha, the two widely studied platforms in the literature on social media and investing.

I scraped 4.70 million tweets from advisers' accounts. Twitter allows users to mention securities by typing a dollar sign before their ticker symbols, e.g. $INTC for the Intel stock, creating searchable symbols called cashtags. Among the extracted tweets, 191354 tweets from 969 unique users contain at least one cashtag. I restrict my sample to tweets mentioning common stocks (share codes 10, 11, and 12) listed on NYSE, NYSE American, or NASDAQ (exchange codes 1, 2, and 3), resulting in 99798 tweets from 697 unique users. These tweets mention 169486 cashtags in total, as each tweet can contain multiple cashtags. Appendix A includes a few examples of the tweets in my data.

I use state-of-the-art machine learning algorithms to create nine sentiment analysis models. I manually label 6143 tweets and train each model to measure the tweet sentiment for each mentioned cashtag as positive, neutral, or negative. To increase the total accuracy of sentiment measurements, I aggregate the outputs of these nine models by taking the most frequent label for each cashtag. I will refer to this aggregate label as the consensus model. The consensus model achieves a total accuracy of 83.1% out-of-sample, which is similar to the figures reported in the literature (Cookson and Niessner (2020)). Appendix B describes all sentiment models and details their performance in the data. Unless otherwise specified, the sentiment in the following analyses is calculated using the consensus model.

2.2. Other Data Sources

I obtain stock data from CRSP for the period 2008 to 2020 corresponding to the time horizon of the Twitter sample. I match each tweet with the first trading day that closes after the tweet's posting time. Therefore, tweets within trading hours match with the same day while after-hour tweets match with the next trading day. I calculate abnormal returns according to the method of Daniel et al. (1997). Following Chen et al. (2014), I aggregate the tweets at the stock/day level before running my tests. My measure of sentiment follows Antweiler and Frank (2004) and is defined for stock i on day t as

$$Sentiment_{i,t} = \log\left(\frac{1 + pos_{i,t}}{1 + neg_{i,t}}\right),$$

where $pos_{i,t}$ and $neg_{i,t}$ are the number of positive and negative tweets about stock i on day t. I also decompose the daily sentiment into positive and negative components defined as

$$Positive_{i,t} = \max\{Sentiment_{i,t}, 0\}$$
$$Negative_{i,t} = \max\{-Sentiment_{i,t}, 0\}.$$

For each stock/day observation, I also calculate the return over the prior week, abnormal turnover, and volatility.

I acquire five-minute intraday prices from TAQ following Holden and Jacobsen (2014). To merge with intraday prices, I shift the posting time of trading-hour tweets to the next five-minute time bin. For example, a tweet posted at 11:42 AM will match the price at 11:45 AM. On the other hand, after-hour tweets are shifted to 9:35 AM on the next trading day to ensure overnight returns do not contaminate my intraday results. In addition to intraday prices, I follow Boehmer et al. (2021) to identify retail trades from TAQ and tag them as buys or sells.

In addition to CRSP and TAQ, I obtain data on analyst recommendations and earnings surprises from IBES, news sentiment from Ravenpack, and adviser characteristics from form ADV submissions. I normalize analyst recommendation changes such that the maximum upgrade (from sell to strong buy) is represented by $+1$. My measure of earnings surprise is the difference between the realized EPS and the analyst consensus forecast normalized by the closing stock price on the day of the earnings announcement. I winsorize this measure at 0.1% and 99.9% to curtail the effect of outliers. Ravenpack provides the sentiment of each news story as an integer between 0 and 100, with 50 representing a neutral story. I map its sentiment measure to the interval $[-1, 1]$. Ravenpack also provides a detailed news taxonomy whose most general level is called a *category*. I aggregate its sentiment at the stock/day/category level and choose the 24 most frequently populated news categories to include in my regressions. I aggregate the sentiment of analyst recommendations, earnings surprises, and news events at the stock/day level before merging them with CRSP. Appendix C describes all variables in detail.

2.3. Summary Statistics

I employ the consensus model for my main analyses throughout this paper. Table 1 describes investment advisers' Twitter activity by year. Overall, 72133 (69%) of tweets are neutral, 36287 (23.1%) are positive, and 12126 (7.9%) are negative. Twitter activity sharply increased in 2011 and continued to grow until 2014, after which it declined for two years and stabilized. Not all advisers tweet every year. On average, 216 advisers have tweeted each year since 2014. On average, advisers mention 1994 stocks per year since 2014. In total, there are 5234 unique stocks in the data.

Because it is difficult to determine the value of neutral tweets, I exclude them from my analyses henceforth. Figure 1 shows the distribution of non-neutral tweets across advisers. In total 274 advisers post non-neutral tweets about 4021 unique stocks. The average (median) adviser posts 191.9 (5) tweets while the standard deviation of the distribution is 1442.6. It is noteworthy that tweeting activity is skewed across advisers. The top five most frequently tweeting advisers post around 76% of all tweets. In section 3.5, I analyze the top five advisers separately.

2.4. Five Facts about Twitter Data

In this section, I review five salient facts about my Twitter data. The first two facts report the cross-sectional distribution of tweets, the third and fourth describe the time series of tweets, and the last fact highlights the distribution of tweeting activity across advisers.

2.4.1. Tweeting Activity is Skewed across Stocks.

Figure 2 shows the time series of positive and negative tweets for six stocks in 2018. Panels (a) and (b) show that Apple and Amazon were tweeted several times per week. On the other hand, panels (c) through (f) illustrate how tweeting frequency rapidly drops across other S&P500 stocks. The 80th percentile of stocks received only six tweets throughout 2018. Therefore, the data shows a high concentration of tweets around a few stocks, while others receive at most a handful of tweets per year. Figure 3 demonstrates how the distribution of tweets flattens after the first ten stocks. Inspired by this observation, I separate the ten most frequently-tweeted stocks in section 3.2. Figure A.2 shows the distribution of tweets across all stocks.

2.4.2. Larger Stocks Receive More Tweets.

Given that larger stocks receive more reports from sell-side analysts (Bhushan (1989)) and social media analysts (Chen et al. (2014)), one might expect advisers to tweet them more often as well. Figure 4 confirms this conjecture. It is also noteworthy that the largest bin of stocks receives disproportionately more tweets. Moreover, the confidence interval is wider for larger bins, indicating that the distribution of tweets grows wider for larger stocks. I compare tweet informativeness for small and large stocks separately in section 3.2.

2.4.3. Tweets are More Likely after News.

Sell-side analysts issue significantly more revisions on days following earnings announcements (Ivković and Jegadeesh (2004)). Therefore, we might expect news to influence tweets as well. Assuming the first market close after the news is on day t, I define the news window as days t and $t+1$. I consider three categories of news:

earnings announcements, analyst revisions, and other news. Figure 6 illustrates how news relates to the frequency of advisers' tweets. Panel (a) shows that 18.4% of all stock/days in CRSP are inside at least one news window. In contrast, panel (b) shows that 58.4% of tweeted stock/days are within a news window. The increase in the probability of tweeting is even more substantial for analyst revisions and earnings announcements, which justifies separating these two categories of news. Given the significant relation between news and tweeting activity, we can ask whether tweets pass through public news or provide novel information. Section 3.3 addresses this question.

2.4.4. Tweets Sentiment is Positively Correlated with Past Returns.

Prior studies have found evidence of return chasing among advisers (Linnainmaa et al. (2021), Mullainathan et al. (2012)). Moreover, Twitter sentiment often reflects recent price movements (Groß-Klußmann et al. (2019)). To investigate the relation between tweeting activity and past returns, I divide all stocks into deciles based on the prior week returns every day and count the number of positive and negative tweets that each decile receives throughout the sample. Figure 5 shows the results. Three patterns stand out. First, the total tweet distribution is J-shaped: the highest and lowest deciles receive more tweets than the middle ones, with the highest decile receiving more than twice as many tweets as any other. Second, the ratio of positive to negative tweets monotonically increases with decile rank. Third, a significant minority of tweets oppose the sentiment of prior returns. For example, the number of positive tweets about the lowest decile is larger than the fifth and sixth deciles. Taken together, patterns two and three imply that tweets partially reflect recent price movements but may contain original information as well. Section 3.4 tests the relationship between prior returns and tweet informativeness.

2.4.5. Tweeting Activity Varies by Adviser Type.

Investment advisers run a variety of businesses.[1] According to data from form ADV filings, they offer services such as managing portfolios, financial planning and consulting, and security pricing and analysis. Therefore, it is reasonable to expect the content of their tweets to depend on the nature of their businesses. To explore this possibility, I identify three types of advisers based on the services they provide:

1. fund managers, who manage portfolios for investment companies or private funds, or otherwise advise private funds;
2. individual managers, who manage portfolios for individual investors or small businesses; and
3. financial planners, who provide financial planning or pension consulting services.

Note that an adviser may belong to more than one category. As such, Figure 7 compares the composition of investment advisers between form ADV data and my sample of tweets. While 67.4% of all advisers are individual managers, they post 96.5% of the tweets. The share of financial planners also increases from 45.3% in panel (a) to 54.4% in panel (b). On the other hand, the percentage of fund managers shrinks from 68.8% in panel (a) to 41.9% in panel (b), and almost all of this decrease comes

[1] The Investment Advisers Act of 1940 defines the term "investment adviser" as any person or firm that engages in the business of providing advice for compensation. For more information, see `https://www.sec.gov/about/offices/oia/oia_investman/rplaze-042012.pdf`.

from fund managers who are neither individual managers nor financial planners. This classification covers 98.4% of investment advisers and 99.3% of their tweets. Inspired by this observation, section 3.5 investigates tweet informativeness for each category of advisers.

3. Methodology and Results

3.1. Predictive Regressions

Are advisers' tweets informative about future stock returns? To formally answer this question, I regress weekly forward abnormal stock returns on my daily measure of tweet sentiment. The regression equation is

$$AbnRet_{i,t+1,t+5} = \alpha + \beta Sentiment_{i,t} + \gamma X_{i,t} + \epsilon_{i,t+1,t+5}, \tag{1}$$

where X represents a vector of control variables. I cluster the standard errors at the stock and month levels following Chen et al. (2014).

Table 2 presents the results of these regressions. Column 1 shows that a one standard deviation increase in tweet sentiment predicts abnormal returns of 6 bps over the next week. Column 2 includes stock controls. The coefficient of tweet sentiment increases to 18 bps, primarily due to past returns. Given that tweets are positively correlated with past returns, this is consistent with the well-established short-term reversal effect (Jegadeesh (1990), Lehmann (1990)). Column 3 adds controls for news events over the prior week. These controls consist of analyst revisions, earnings surprises, and the sentiment of 24 news categories from Ravenpack. The coefficient of tweet sentiment decreases to 12 bps, consistent with the fact that some tweets reflect the sentiment of recent news events. The standard deviation of sentiment is

0.62, which is slightly less than the sentiment of a stock/day with a single positive tweet ($log(2) \approx 0.69$). Therefore, a single positive tweet predicts 13.4 bps weekly abnormal returns. Finally, column 4 decomposes the sentiment into positive and negative components. Both components predict significant abnormal returns with similar magnitudes and correct signs. These results are robust to dropping neutral stock/days, using each of the other nine sentiment algorithms, adjusting returns using Fama-French factor models, and alternative return horizons from one day to three months. Appendix D provides the results of such alternative specifications.

Advisers might use their tweets to run pump-and-dump schemes or generate uninformative temporary price pressure. If so, the price of tweeted stocks should be hump-shaped after the tweets. In other words, the stock would experience a price run-up, and then slide when advisers wind down their portfolios. In contrast, Figure 8 shows abnormal returns do not reverse over the next three months. On the contrary, the difference between abnormal returns following positive and negative stock/days grows to 1.3% over the next three months. This result undermines the pump-and-dump hypothesis by showing that, on average, advisers would lose money on their schemes.

Return measurement starts from the day after the tweet in equation 1. Thus, the point estimates in Table 2 do not include intraday returns. Figure 9 shows that, on average, prices increase (decrease) by 10 (22) bps after positive (negative) tweets and before the market closes on the same trading day. Therefore, advisers' tweets predict intraday returns as well. This result implies that the estimates of Table 2 are conservative measures of the true predictive power of advisers' tweets.

3.2. Cross-Sectional Heterogeneity

As shown in section 2.4, a few stocks receive tweets at least every few days while others get a few tweets per year. Given that the most frequently-tweeted stocks are often household names, their tweets are more likely to be part of uninformed conversations. As such, I hypothesize that tweets about such stocks are less informative. Furthermore, larger stocks have more transparent information environments and are more liquid. As a result, their prices reflect fundamentals more accurately. Therefore, my second hypothesis is that tweets about larger stocks are also less informative.

To test these hypotheses, I consider three subsamples of stocks: (1) the top 10 most frequently-tweeted stocks, (2) S&P500 stocks not in the top 10, and (3) stocks not included in the first two categories. Because the S&P500 index usually includes the largest stocks, the third category includes smaller ones. I interact the sentiment with indicators for the first two categories, taking other stocks as the baseline group. The regression equation is

$$
\begin{aligned}
AbnRet_{i,t+1,t+5} = \alpha &+ \beta_1 Sentiment_{i,t} \\
&+ \beta_2 Sentiment_{i,t} \times \mathbb{1}\{Top10\}_i + \beta_3 Sentiment_{i,t} \times \mathbb{1}\{S\&P500\}_{i,t} \\
&+ \beta_4 \mathbb{1}\{Top10\}_i + \beta_5 \mathbb{1}\{S\&P500\}_{i,t} + \gamma X_{i,t} + \epsilon_{i,t+1,t+5}
\end{aligned}
$$

Table 3 reports the results. The coefficient of sentiment for other stocks is 20 bps, which is larger than the coefficient in column 3 of Table 2 and statistically significant. On the other hand, tweet sentiment is considerably less informative for the first category. Even though the interaction term is insignificant, the total coefficient for sentiment in the top 10 group is only 2 bps (20 bps - 18 bps), which is economically small and statistically insignificant (p-Value=88.4%). Similar results obtain for the S&P500 stocks. The coefficient on the interaction term is -17 bps and significant. The total coefficient of sentiment for the S&P500 stocks is 3 bps (p-Value=46.2%). Therefore, I find evidence for both hypotheses discussed above. It is also noteworthy that the top 10 stocks have performed well during my sample period. Therefore, the coefficient for the top 10 indicator is large (29 bps) and significantly different from zero.

3.3. Tweets and News

After establishing that advisers' tweets predict returns, I inspect the source of their information. Inspired by the literature on analyst recommendations (Li et al. (2015)), I consider two channels. On the one hand, investment advisers can process public information. Under this hypothesis, advisers post tweets around corporate news and their tweets correlate with the market reaction to the news. On the other hand, advisers can post informative tweets on days without news, thereby disseminating nonpublic information. Henceforth, I refer to these channels as news processing and novel information. To gauge how news is reflected in advisors' tweets, I first ask how often stocks are tweeted based on the type of news on that day and the recent past. I measure tweeting activity with two variables: a dummy for whether stock i was tweeted with a non-neutral sentiment on day t, and another for one plus log number of non-neutral tweets about stock i on day t. Following the analysis in section 2.4.3, I separate analyst revisions and earnings announcements from other news categories.

To capture the effect of news on tweeting activity on the extensive margin, I run regressions of the form

$$y_{i,t} = \alpha + \beta_1 RecDay_{i,t} + \beta_2 EarnDay_{i,t} + \beta_3 OtherNewsDay_{i,t} + \gamma Y_{i,t} + \eta_i + \epsilon_{i,t},$$

where $RecDay$, $EarnDay$, and $OtherNewsDay$ are dummy variables indicating whether there was news of that category on day t or $t-1$. The dependent variable can be one of the variables described above. $Y_{i,t}$ is a vector containing five lags of the dependent variable. On the intensive margin, I break down the news days even further based on the severity and direction of the news. To do so, I replace each of the independent variables in the above regression with dummy variables indicating the tercile of the event sentiment. The highest (lowest) tercile represents the most positive (negative) event day.

Table 4 reports the results of these regressions. Column 1 demonstrates that tweets are more likely on days with news. Analyst revisions and earnings announcements increase the probability of tweeting by 4.14% and 1.76%, respectively. In contrast, other news categories increase the probability of tweeting by only 0.18%. Column 2 shows that news days receive more tweets on average. Similar to column 1, analyst revisions and earnings announcements are more correlated with the number of tweets. Columns 3 and 4 show that the association between news and tweeting activity is not monotonic in the severity or sentiment of news. Nevertheless, the probability of tweeting increases for all categories. In short, Table 4 suggests that tweeting activity is higher on news days. These results are robust to including more lags of the main variables or changing the length of the window for calculating the independent variables.

Advisers may provide profitable signals through confirming or contradicting the prevailing news sentiment.[2] Hence, it is reasonable to ask how often tweets agree with the sentiment of news events. To address this question, I divide each category of news into positive and negative within the subsample of tweeted news days. Within each sentiment subcategory, I calculate the share of tweeted days with the same and the opposite sentiment. Table 5 presents the results of this categorization. Even though tweets are more likely to agree with news, a substantial fraction of them contradict news sentiment. Disagreement between news and tweets is smallest for the analyst revisions (8.84% of tweeted revisions) but increases for earnings announcements (29.83%) and other news (43.63%). Hence there is some evidence of advisers both confirming and contradicting the prevailing news sentiment.

Finally, I formally test the two hypotheses discussed at the beginning of this section by interacting tweet sentiment with indicators for news days. The regression equation is

$$\begin{aligned} AbnRet_{i,t+1,t+5} = \alpha &+ \beta_1 Sentiment_{i,t} + \beta_2 Sentiment_{i,t} \times RecDay_{i,t} + \\ &\beta_3 Sentiment_{i,t} \times EarnDay_{i,t} + \beta_4 Sentiment_{i,t} \times OtherNewsDay_{i,t} \\ &+ \beta_5 RecDay_{i,t} + \beta_6 EarnDay_{i,t} + \beta_7 OtherNewsDay_{i,t} \\ &+ \gamma X_{i,t} + \epsilon_{i,t+1,t+5}, \end{aligned}$$

where *RecDay*, *EarnDay*, and *OtherNewsDay* are dummy variables as defined in equation 3.3. X is a vector of control variables containing prior week returns,

[2] Even if tweets never disagree with news, they can still provide a signal through selectively confirming them.

volatility of returns over the prior month, and abnormal turnover. In this regression, $\beta_1 > 0$ implies that tweets provide information on non-news days. In addition, $\beta_1 + \beta_2 > 0$ means tweets provide information by interpreting analyst revisions. Similarly, we can test whether tweets interpret earnings announcements ($\beta_1+\beta_3 > 0$) and other news ($\beta_1 + \beta_4 > 0$).

Table 6 provides evidence supporting both news processing and novel information channels. On the one hand, the estimate for tweet sentiment is positive and significant across all specifications, thereby supporting the novel information channel. A one standard deviation increase in sentiment on a non-news day predicts 10 to 14 bps of abnormal returns over the next week. On the other hand, in column 4, $\beta_1 + \beta_2 = 0.08$ (p-Value = 8.1%) and $\beta_1 + \beta_3 = 0.25$ (p-Value = 0.5%), supporting the news processing channel for analyst revisions and earnings announcements. The evidence is weaker on other news days: $\beta_1 + \beta_4 = 0.09$ with a p-Value of 12%. Given that advisers post fewer tweets on other news days, the higher measurement error, and hence higher p-Value, is not surprising.

3.4. Curating Past Winners and Losers

Previous studies find evidence of return chasing among investment advisers (Linnainmaa et al. (2021), Mullainathan et al. (2012)). Given consistent patterns discussed in section 2.4, I ask how the correlation between past returns and tweeting activity affects tweet informativeness. If advisers only chase returns, the sentiment of their tweets will not provide a profitable signal after large price movements. Indeed, tweet sentiment would negatively predict future returns due to the short-term reversal effect (Jegadeesh (1990), Lehmann (1990)). On the other hand, stocks receive more attention after large price movements (Barber and Odean (2007)). Hence, advisers might selectively tweet about stocks after large returns to highlight their stock-picking skills when their pick receives more attention. This hypothesis implies that such tweets could be informative about future returns.

To investigate the relationship between tweet informativeness and past returns further, I test the predictive power of tweets in subsamples based on deciles of past returns. Table 7 shows that for the highest and the lowest deciles, the coefficient on tweets sentiment is large (31 bps for the lowest and 17 bps for the highest decile) and significant. This evidence is consistent with the attention timing hypothesis. The point estimate is positive but insignificant for deciles 2, 3, 4, and 9 and negative but insignificant for deciles 5 through 8.

To formally test my hypotheses, I interact tweet sentiment with an indicator for extreme (the highest and lowest) deciles of past returns. Table 8 reports the results, varying the horizon of past returns from one day to one week. Across the three columns, the coefficient on the sentiment is small and insignificant. Thus, on average tweet sentiment is uninformative for stocks outside of the extreme deciles. On the other hand, the coefficient on the interaction term is 31 to 34 bps and significant at 1%. Consistent with the attention timing hypothesis, these results suggest that advisers curate recent winners and losers and tweet about stocks that continue to outperform in the future.

3.5. Adviser Heterogeneity and Tweet Informativeness

Given the heterogeneity in the businesses registered as investment advisers, it is reasonable to expect heterogeneity in the informativeness of their tweets. On the one hand, fund managers are often directly compensated for their stock-picking skills. Therefore, they might be better informed about stocks than other advisers. Indeed, Chen et al. (2017) find that nearly two-thirds of hedge fund managers have positive alpha. On the other hand, fund managers might also conceal their information to reduce the risk of front-running. In addition, Gurun et al. (2018) show that in the aftermath of the Madoff scandal, retail clients withdrew fewer assets from planners, suggesting that financial planning might build a trust relationship between customers and their advisers. Hence, financial planners might reciprocate their clients' trust by providing them with useful information about stocks. In line with this point, financial planners tweet a substantial amount of other useful information, such as best practises for retirement planning and tips for saving on taxes. Thus, whose tweets are more informative remains an empirical question.

To answer this question, I analyze fund managers and planners separately. Following the methodology in section 2.2, I create three sentiment measures using tweets from fund managers, financial planners, and advisers who do not belong in either category. In column 1 of Table 9, I repeat my main regression after including all three sentiment measures. A one standard deviation increase in financial planners' sentiment predicts abnormal returns of 12 bps, which is statistically significant. On the other hand, the coefficients of sentiment for fund managers and other advisers are 6 and 9 bps, both insignificant. In column 2, I separate the top five frequently-tweeting advisers from each group. The less active financial planners post more informative tweets than financial planners in the top five. The difference (30 bps) is economically large but statistically insignificant. On the other hand, fund managers who tweet frequently post more informative tweets compared to other fund managers. In fact, the tweet sentiment for the most frequently-tweeting fund managers predicts abnormal returns with a significant coefficient of 11 bps. The difference term has a large (-28 bps) though statistically insignificant coefficient. The results in Table 9 indicate that financial planners' tweets are more informative than other investment advisers, and that tweet informativeness is not restricted to the most active advisers.

3.6. Advisers' Tweets and Retail Investors

Prior literature finds that retail investors react to online stock analyses (Farrell et al. (2022)). Given that financial planners tend to serve retail clients, it is natural to ask whether retail investors react to their tweets. Therefore, I regress retail investors' order imbalance over the next trading week on the sentiment of the tweets. The regression equation is

$$RetailOIB_{i,t+1,t+5} = \beta Sentiment_{i,t} + \gamma X_{i,t} + \eta_i + \xi_t + \epsilon_{i,t+1,t+5}. \tag{2}$$

In addition to all the control variables in regression 1, I control for $RetailOIB_{i,t-4,t}$ to account for the autocorrelation in retail order imbalance. Also, the relative timing of tweets and other variables follows the same convention as in Table 2. Because the measure of retail order imbalance in Boehmer et al. (2021) is only valid after 2010, I exclude 2008-2009 from my analysis in this section.

Table 10 reports the results of these regressions. Column 1 shows that a one-standard deviation increase in tweet sentiment increases retail order imbalance by 2.3% of a standard deviation. I consider the magnitude reasonable given the reach of financial advisers' tweets. Columns 2 and 3 show that day fixed effects and controls have small effects on the estimate. Column 4 shows that both positive and negative stock/days predict changes in the retail order imbalance with the appropriate signs. However, the coefficient for positive tweets (0.029) is almost twice as large as that of negative tweets (-0.015). Because the measure of Boehmer et al. (2021) is more accurate before 2015, I restrict my sample to 2010-2015 in column 5 and find that the point estimate remains almost the same.

Advisers' tweets predict retail order imbalance at the intraday frequency as well. Figure 10 shows the difference in retail order imbalances within a trading day around positive and negative tweets. The curve is near zero before tweets but increases to 0.05 at the half-hour interval when the tweet is posted and remains elevated for the next 5 hours.

Taken together, Table 10 and Figure 10 show that tweet sentiment can predict retail order imbalance. This predictive power may be causal. In other words, it is possible that advisers' tweets prompt some retail investors to trade. On the other hand, the predictive power of tweets for retail order imbalance could be endogenous. For example, advisers might time their tweets to overlap with when retail investors are trading in a certain stock, or advisers and retail investors could be reacting to the same event (e.g. news). I remain agnostic as to the relative contribution of each channel. However, given that advisers *choose* to post informative tweets about stocks, both hypotheses imply that they serve as information intermediaries for retail investors, regardless of whether retail investors trade on that information or not.

Boehmer et al. (2021) also report that retail order imbalance predicts abnormal returns. In their Fama-MacBeth (1973) regressions, the difference between next-week abnormal returns for stocks in the 75th and 25th percentiles of retail order imbalance is around 11 bps. It is reasonable to ask whether retail order imbalance can explain return predictability in my setup. I bring three arguments against this hypothesis. First, in their setup, abnormal returns *follow* an increase in retail order imbalance, while in my setup they are concurrent. Second, Table V in their paper reports that the predictive power of retail order imbalance decreases over time and becomes insignificant at 12 weeks. In contrast, Figure 8 shows that the predictive power of tweets grows over time and is still significant in 12 weeks. Third, in untabulated results, I control for past retail order imbalance in my regressions and find that my main results do not change. I conclude that advisers' tweets predict abnormal returns even after accounting for retail order imbalance.

4. Conclusion

Prior literature has tested whether Twitter sentiment contains information about stock prices with inconsistent results (Bartov et al. (2018), Giannini et al. (2018), Ballinari and Behrendt (2021)). I focus on a group of professionals on Twitter, namely investment advisers and show that a one standard deviation increase in the sentiment of their tweets predicts 12 bps of abnormal returns over the next trading week. Regarding the source of advisers' information, my evidence suggests that they process public news and disclose novel information. Among investment advisers, those offering financial planning services post more informative tweets. In addition, retail investors trade in the direction of tweets over the next week.

References

Antweiler, Werner, and Murray Z Frank, 2004, Is all that talk just noise? the information content of internet stock message boards, *The Journal of finance* 59, 1259–1294.

Azimi, Mehran, and Anup Agrawal, 2021, Is positive sentiment in corporate annual reports informative? evidence from deep learning, *The Review of Asset Pricing Studies* 11, 762–805.

Ballinari, Daniele, and Simon Behrendt, 2021, How to gauge investor behavior? a comparison of online investor sentiment measures, *Digital Finance* 3, 169–204.

Barber, Brad M., and Terrance Odean, 2007, All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors, *The Review of Financial Studies* 21, 785–818.

Bartov, Eli, Lucile Faurel, and Partha S Mohanram, 2018, Can twitter help predict firm-level earnings and stock returns?, *The Accounting Review* 93, 25–57.

Bergstresser, Daniel, John MR Chalmers, and Peter Tufano, 2008, Assessing the costs and benefits of brokers in the mutual fund industry, *The Review of Financial Studies* 22, 4129–4156.

Bhushan, Ravi, 1989, Firm characteristics and analyst following, *Journal of accounting and economics* 11, 255–274.

Boehmer, Ekkehart, Charles M Jones, Xiaoyan Zhang, and Xinran Zhang, 2021, Tracking retail investor activity, *The Journal of Finance* 76, 2249–2305.

Campbell, John L, Jenna D'Adduzio, and James Moon, 2021, Everyone has an opinion: The informativeness of social media's response to management guidance, *Working Paper* .

Campbell, John L, Matthew D DeAngelis, and James R Moon, 2019, Skin in the game: Personal stock holdings and investors' response to stock analysis on social media, *Review of Accounting Studies* 24, 731–779.

Chalmers, John, and Jonathan Reuter, 2020, Is conflicted investment advice better than no advice?, *Journal of Financial Economics* 138, 366–387.

Chen, Hailiang, Prabuddha De, Yu Jeffrey Hu, and Byoung-Hyoun Hwang, 2014, Wisdom of crowds: The value of stock opinions transmitted through social media, *The Review of Financial Studies* 27, 1367–1403.

Chen, Yong, Michael Cliff, and Haibei Zhao, 2017, Hedge funds: The good, the bad, and the lucky, *Journal of Financial and Quantitative Analysis* 52, 1081–1109.

Cookson, J Anthony, Joseph Engelberg, and William Mullins, 2021, Echo chambers, *Working Paper* .

Cookson, J Anthony, and Marina Niessner, 2020, Why don't we agree? evidence from a social network of investors, *The Journal of Finance* 75, 173–228.

Crane, Alan, and Kevin Crotty, 2020, How skilled are security analysts?, *The Journal of Finance* 75, 1629–1675.

Crawford, Steven, Wesley Gray, Bryan R Johnson, and Richard A Price III, 2018, What motivates buy-side analysts to share recommendations online?, *Management science* 64, 2574–2589.

Crawford, Steven S, Wesley R Gray, and Andrew E Kern, 2017, Why do fund managers identify and share profitable ideas?, *Journal of Financial and Quantitative Analysis* 52, 1903–1926.

Dangl, Thomas, and Stefan Salbrechter, 2021, News sentiment and equity returns, *Working Paper* .

Daniel, Kent, Mark Grinblatt, Sheridan Titman, and Russ Wermers, 1997, Measuring mutual fund performance with characteristic-based benchmarks, *The Journal of finance* 52, 1035–1058.

Davis, James L, Eugene F Fama, and Kenneth R French, 2000, Characteristics, covariances, and average returns: 1929 to 1997, *The Journal of Finance* 55, 389–406.

De la Pera, Daniela, 2021, Disclosure softness of corporate language, *Working Paper* .

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova, 2018, Bert: Pre-training of deep bidirectional transformers for language understanding, *arXiv preprint arXiv:1810.04805* .

Dim, Chukwuma, 2020, Should retail investors listen to social media analysts? evidence from text-implied beliefs, *Working Paper* .

Fama, Eugene F, and Kenneth R French, 2015, A five-factor asset pricing model, *Journal of financial economics* 116, 1–22.

Fama, Eugene F, and James D MacBeth, 1973, Risk, return, and equilibrium: Empirical tests, *Journal of political economy* 81, 607–636.

Farrell, Michael, T Clifton Green, Russell Jame, and Stanimir Markov, 2022, The democratization of investment research and the informativeness of retail investor trading, *Journal of Financial Economics* 145, 616–641.

Foerster, Stephen, Juhani T Linnainmaa, Brian T Melzer, and Alessandro Previtero, 2017, Retail financial advice: does one size fit all?, *The Journal of Finance* 72, 1441–1482.

Giannini, Robert, Paul Irvine, and Tao Shu, 2018, Nonlocal disadvantage: An examination of social media sentiment, *The Review of Asset Pricing Studies* 8, 293–336.

Groß-Klußmann, Axel, Stephan König, and Markus Ebner, 2019, Buzzwords build momentum: Global financial twitter sentiment and the aggregate stock market, *Expert Systems with Applications* 136, 171–186.

Gurun, Umit G, Noah Stoffman, and Scott E Yonker, 2018, Trust busting: The effect of fraud on investor behavior, *The Review of Financial Studies* 31, 1341–1376.

Hackethal, Andreas, Roman Inderst, and Steffen Meyer, 2010, Trading on advice .

Holden, Craig W, and Stacey Jacobsen, 2014, Liquidity measurement problems in fast, competitive markets: Expensive and cheap solutions, *The Journal of Finance* 69, 1747–1785.

Ivković, Zoran, and Narasimhan Jegadeesh, 2004, The timing and value of forecast and recommendation revisions, *Journal of Financial Economics* 73, 433–463.

Jegadeesh, Narasimhan, 1990, Evidence of predictable behavior of security returns, *The Journal of finance* 45, 881–898.

Lehmann, Bruce N, 1990, Fads, martingales, and market efficiency, *The Quarterly Journal of Economics* 105, 1–28.

Li, Edward Xuejun, K Ramesh, Min Shen, and Joanna Shuang Wu, 2015, Do analyst stock recommendations piggyback on recent corporate news? an analysis of regular-hour and after-hours revisions, *Journal of Accounting Research* 53, 821–861.

Linnainmaa, Juhani T, Brian T Melzer, and Alessandro Previtero, 2021, The misguided beliefs of financial advisors, *The Journal of Finance* 76, 587–621.

Loh, Roger K., and René M. Stulz, 2010, When Are Analyst Recommendation Changes Influential?, *The Review of Financial Studies* 24, 593–627.

Loughran, Tim, and Bill McDonald, 2016, Textual analysis in accounting and finance: A survey, *Journal of Accounting Research* 54, 1187–1230.

Mullainathan, Sendhil, Markus Noeth, and Antoinette Schoar, 2012, The market for financial advice: An audit study, *NBER Working Paper* .

Rothman, Denis, 2021, *Transformers for Natural Language Processing: Build innovative deep neural network architectures for NLP with Python, PyTorch, TensorFlow, BERT, RoBERTa, and more* (Packt Publishing Ltd).

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin, 2017, Attention is all you need, *Advances in neural information processing systems* 30.

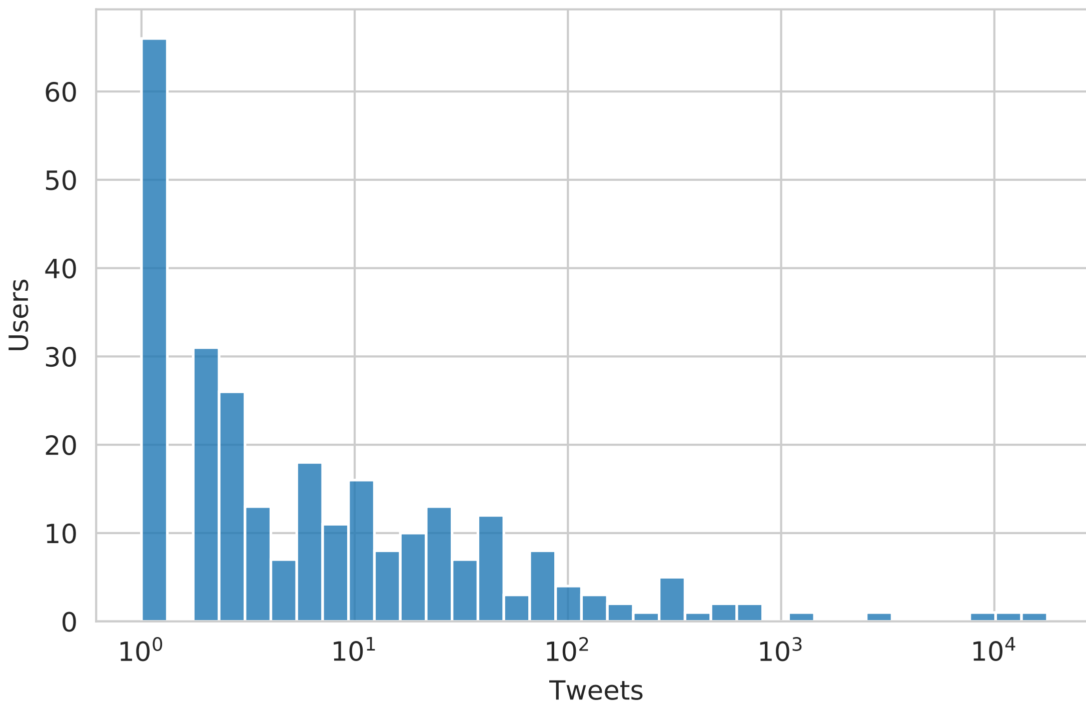


Figure 1: The Distribution of Tweets among Users

This figure shows a histogram of the number of non-neutral stock tweets per adviser. The sample spans 2008-2020. Sentiments were assigned using the consensus algorithm.

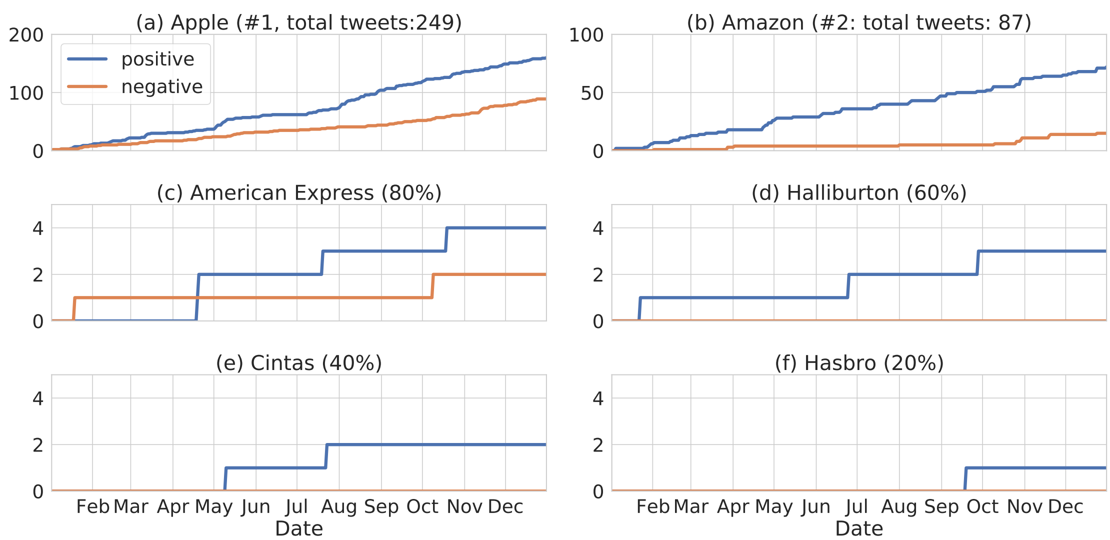


Figure 2: Time Series of Tweets for Six Stocks

This figure shows the cumulative number of positive and negative tweets in 2018 about six stocks. Apple and Amazon were the two most frequently tweeted stocks in year. American Express, Halliburton, Cintas, and Hasbro represent quintiles of nonneutral-tweet count across S&P500 stocks in 2018.

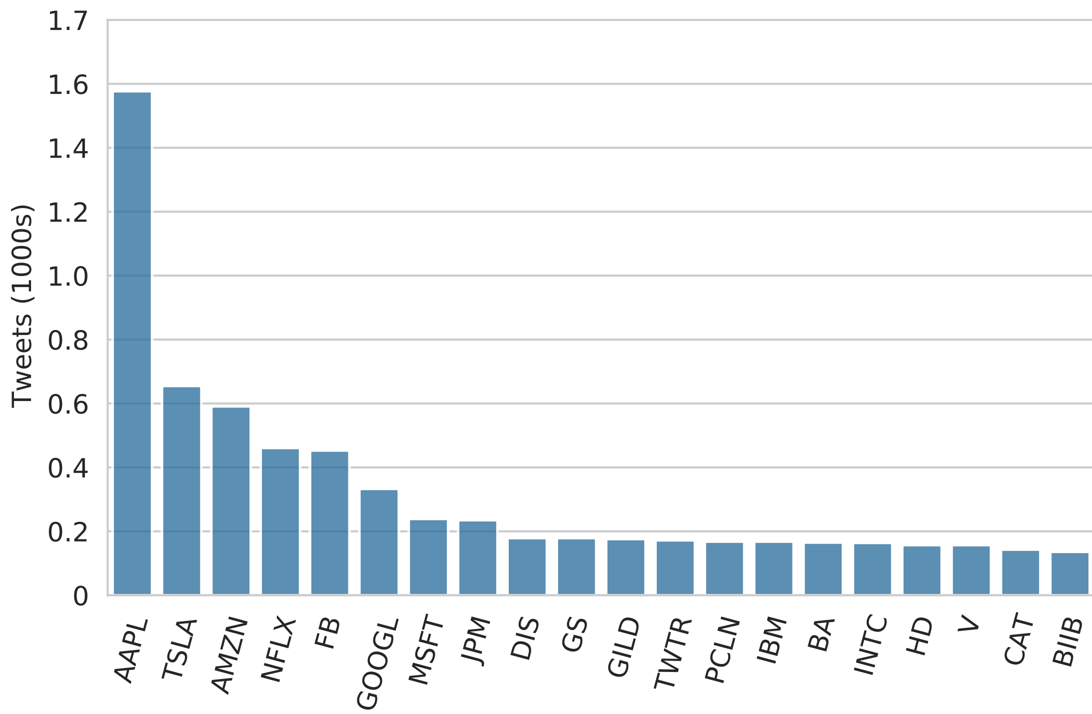


Figure 3: The 20 Most Frequently Tweeted Stocks

This figure shows the number of non-neutral tweets for the 20 most frequently tweeted stocks. I exclude neutral tweets before ranking the stocks.

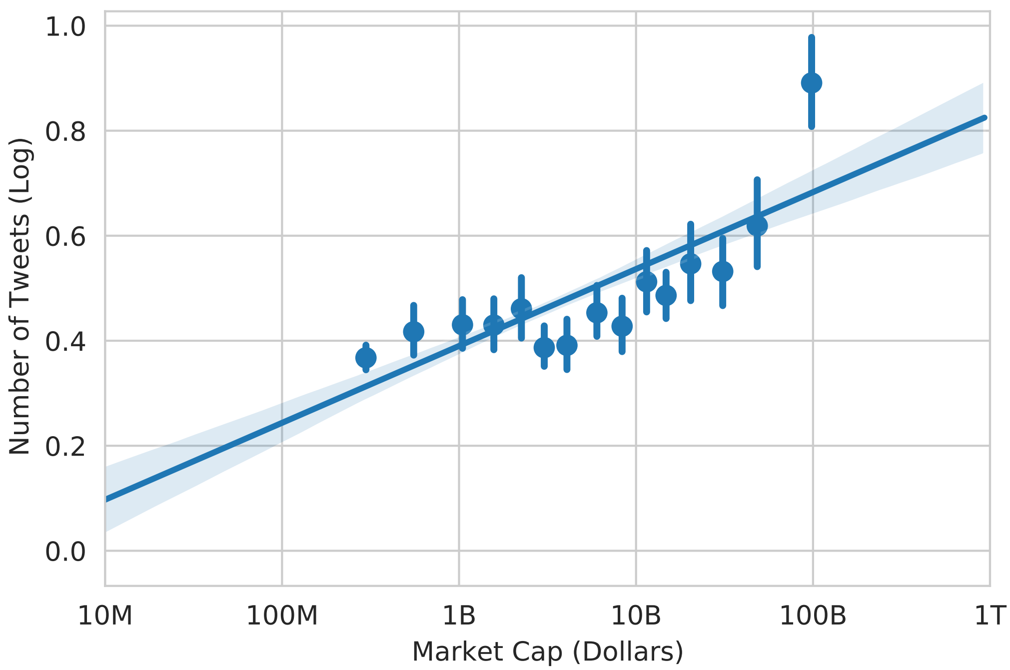


Figure 4: Tweeting Activity and Size

This figure shows a binscatter plot of the log number of nonneutral tweets against market capitalization for stocks tweeted in 2018. The dots and line segments represent the means and 95% confidence intervals of bins. The line and shaded area display the best linear fit and its 95% confidence area.

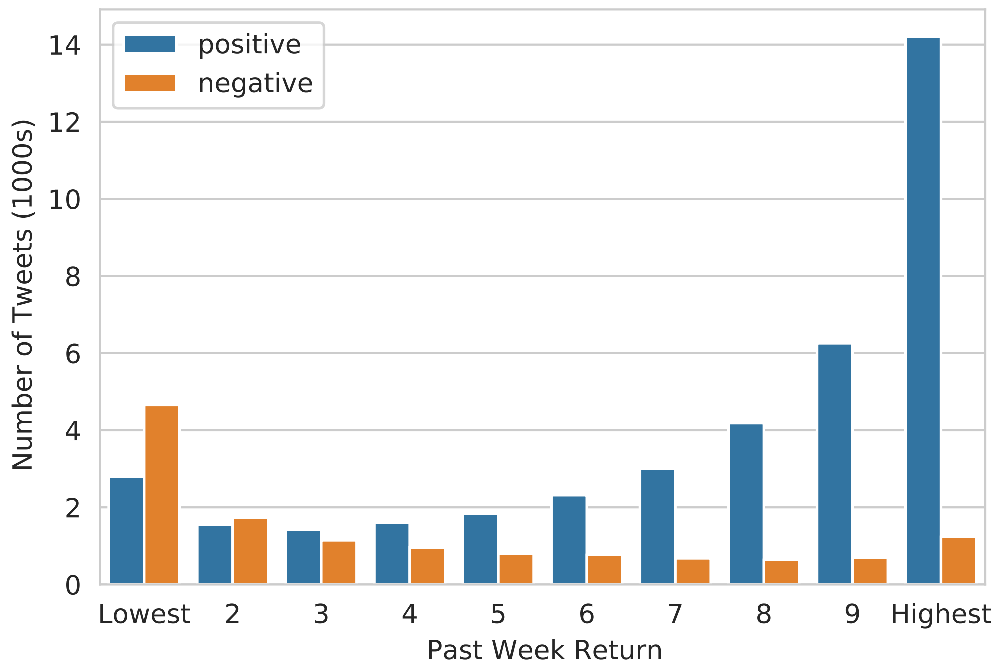
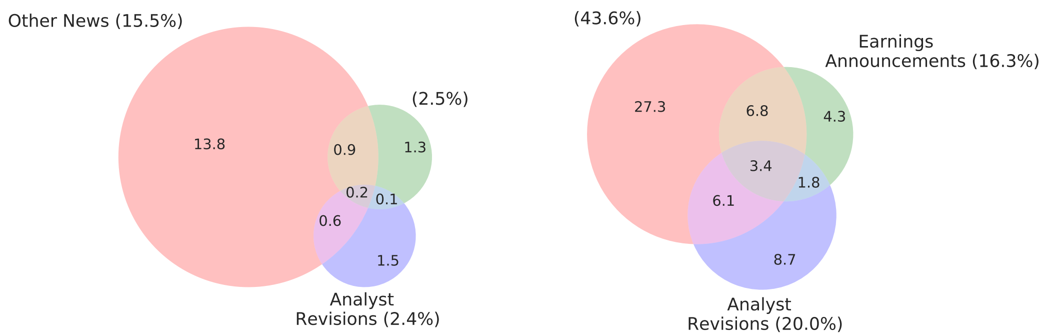


Figure 5: Tweeting Activity and Past Return

This figure shows the distribution of positive and negative tweets across deciles of return over the week prior to the tweet. Decile breakpoints are calculated using only NYSE stocks.



(a) All Stock/Days (Total: 18.4%) (b) Only Tweeted Stock/Days (Total: 58.4%)

Figure 6: Share of News Days

Defining news period as $[t, t+1]$ where t is the day a news story is released, this figure shows a breakdown of days based on types of news periods they are included in. (a) The universal set is all CRSP stock/days from 2008 to 2020 (b) The universal set is stock/days when advisers tweet. The numbers indicate the percentage of observations in each subset. The *total* numbers indicate the percentage of observations belonging to at least one type.

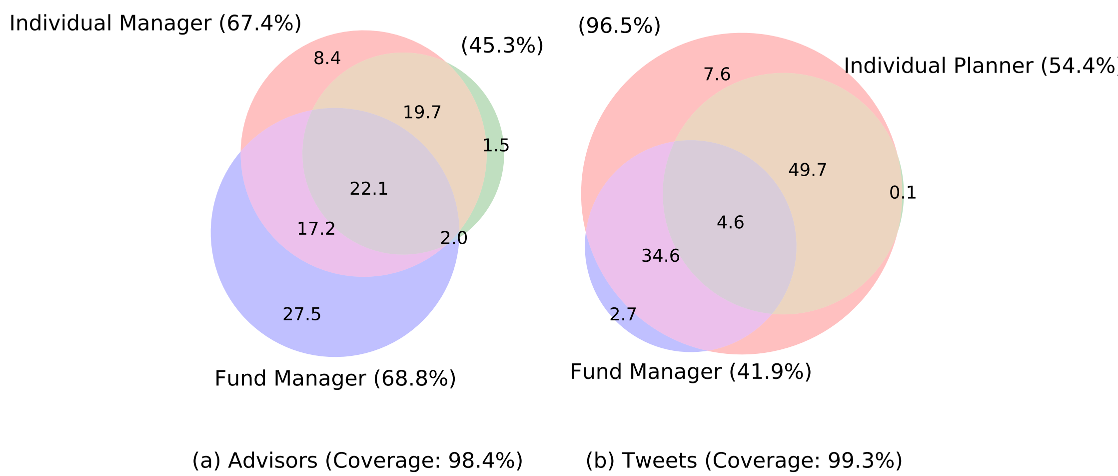


Figure 7: The Composition of Advisers

This figure shows Venn diagrams of the types of advisers in (a) a panel of all registered advisers from 2008 to 2020, and (b) in advisers' tweets. The numbers indicate the percentage of observations in each subset. The *total* numbers indicate the percentage of observations belonging to at least one type.

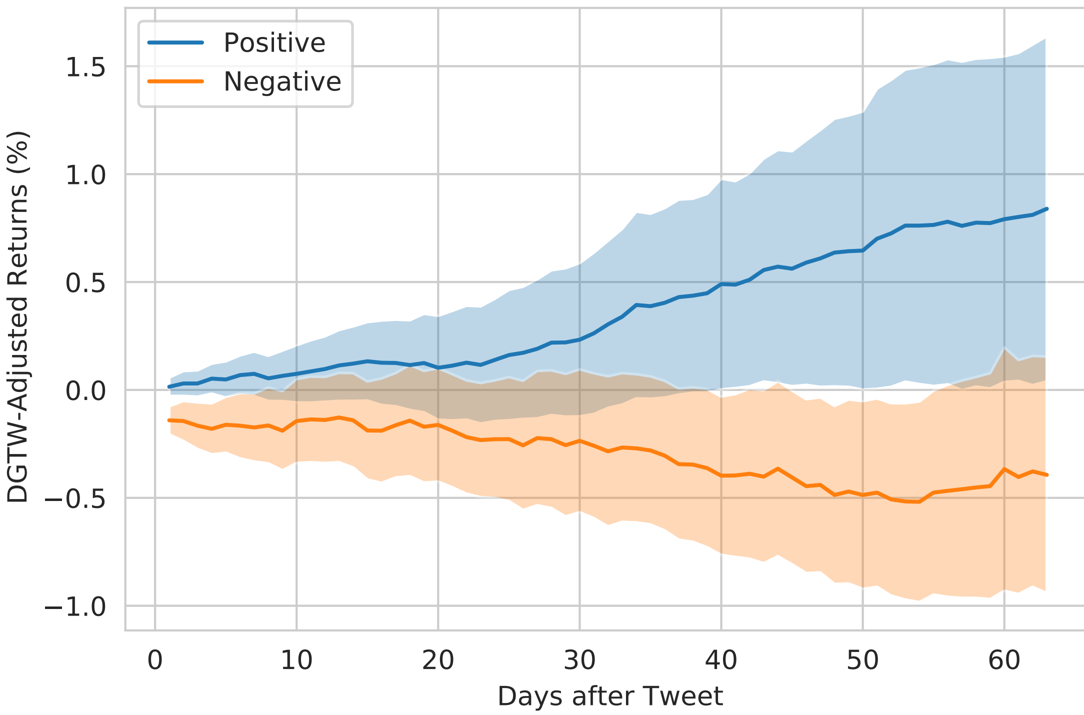


Figure 8: Time Series of Abnormal Returns

This figure shows the time series of abnormal returns up to three months (63 trading days) after stock/days with positive and negative sentiment. Returns are adjusted following Daniel et al. (1997). The shaded areas represent 95% confidence intervals.

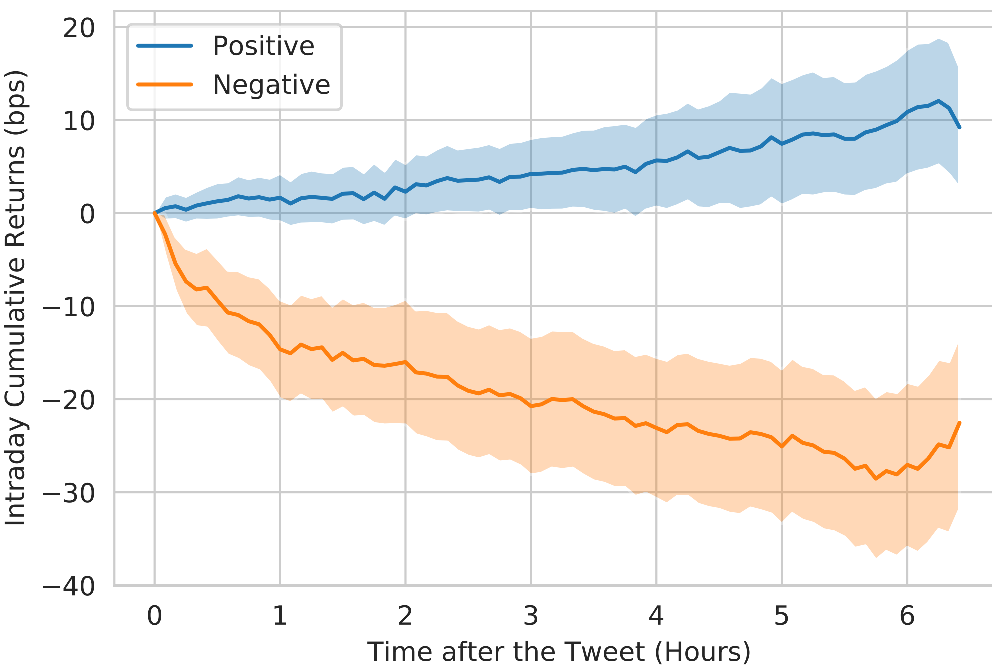


Figure 9: Intraday Returns after Tweets

This figure shows cumulative five-minute returns following positive and negative tweets. For trading-hour tweets, time zero is the first five-minute bin after the tweets' posting time. For after-hour tweets, time zero is 9:35 AM on the next trading day. I drop tweets at the first close after time zero. Shaded areas represent 95% confidence intervals.

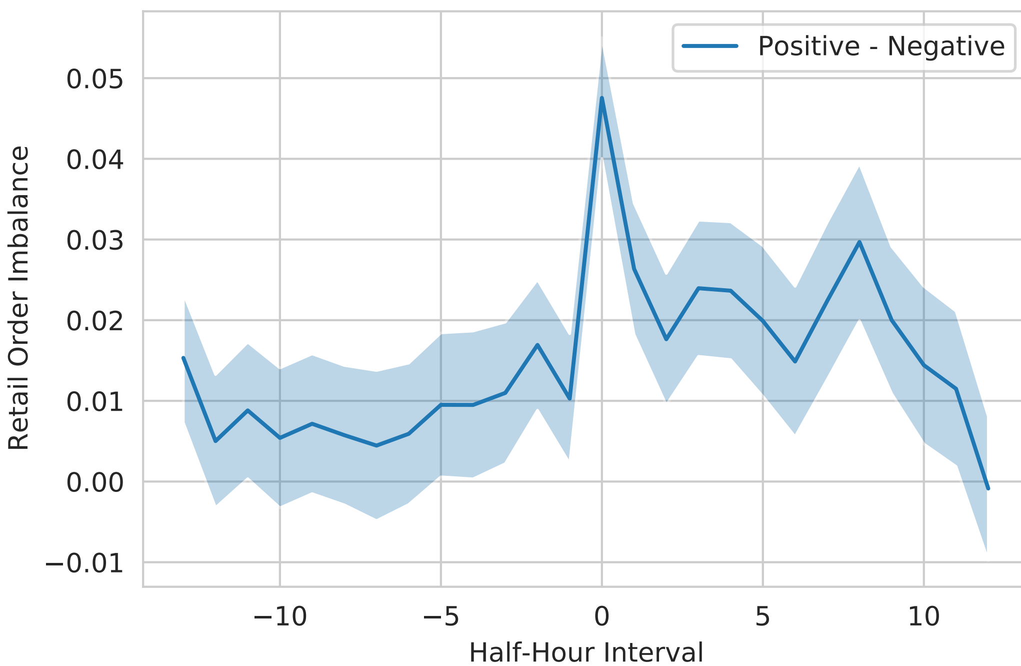


Figure 10: Retail Order Imbalance around Tweets

This figure shows the difference between retail order imbalance for half-hour intervals around positive and negative tweets. The shaded are represents the 95% confidence interval. Time zero represents the first time bin ending after the tweet. In particular, for after-hour tweets it represents 9:30 AM to 10:00 AM on the next trading day.

Table 1: Tweeting Activity by Year

This table describes advisers' tweeting activity by year. The first column shows the number of advisers who tweeted. The second column shows the number of stocks tweeted. Columns 3-5 break down the number of tweets by sentiment. Columns 6-9 break down the tweeted stock/days by sentiment. The sentiment of a stock/day is positive (negative) if the number of positive tweets is strictly more (less) than the number of negative tweets for that stock/day. The last row shows the total value of each variable in 2008-2020. Sentiments are measured with the consensus algorithm.

			Tweets			Stock/Days		
	Advisers	Stocks	Negative	Neutral	Positive	Negative	Neutral	Positive
2008	1	155	2	375	30	2	340	30
2009	6	285	1	512	51	1	479	51
2010	19	365	13	519	127	12	472	117
2011	52	1931	1456	3624	2997	1410	2775	2881
2012	109	2229	3027	5072	4639	2874	3264	4380
2013	149	2298	2565	6313	6042	2404	3901	5633
2014	195	2655	2262	14044	8790	2063	8907	8109
2015	206	2398	1465	16328	4347	1334	10807	4032
2016	221	1787	613	13676	2569	552	8855	2370
2017	216	1659	362	15175	2127	287	9166	1923
2018	226	1688	634	12717	2182	521	7656	1974
2019	209	1441	565	12802	2732	402	6360	2475
2020	237	2328	368	15750	2613	264	9151	2312
Total	697	5234	13333	116907	39246	12126	72133	36287

Table 2: Regressions of Abnormal Returns on Tweet Sentiment

This table reports the results of the regression

$$AbnRet_{i,t+1,t+5} = \alpha + \beta Sentiment_{i,t} + \gamma X_{i,t} + \epsilon_{i,t+1,t+5},$$

on a panel of CRSP stocks spanning from January 1, 2008 to December 31, 2020. *AbnRet* are calculated using the method of Daniel et al. (1997) and in percentage points. *Sentiment* is defined as the natural log of the ratio of one plus the number of positive tweets to one plus the number of negative tweets. *Positive* is defined as the maximum of sentiment and zero. *Negative* is the maximum of -1 times sentiment and zero. These three variables are normalized by the standard deviation of Sentiment on days when it is non-zero. *Prior Week Return* is the stock return over the period $[t-4, t]$. *Abnormal Turnover* is the difference between the ratio of the number of shares traded to total shares outstanding at day t and the average of the same ratio over the six-month period ending one month prior to that day. *Volatility* is the standard deviation of daily returns over the month leading to day t. *Analyst Revision* is the rolling sum over the prior two days of the analyst revisions normalized such that a value of 1 represents an upgrade from strong sell to strong buy. *Earnings Surprise* is the sum over the prior two days of the difference between realized quarterly EPS and IBES analyst consensus forecast normalized to stock price on the announcement day. *News Sentiment* represents 24 control variables, each equal to the rolling sum over the prior two days of the Ravenpack sentiment for the a news category conditional on a novelty and relevance score of at least 75. Each Ravenpack sentiment observation is scaled to the interval $[-1, 1]$. Appendix C includes a list of news categories included in this table. Standard errors are double-clustered at the stock and month levels. ***, **, and * represent statistical significance at 1%, 5%, and 10% respectively.

	(1)	(2)	(3)	(4)
Sentiment	0.06*	0.18***	0.12***	
	(0.03)	(0.04)	(0.04)	
Positive				0.12***
				(0.05)
Negative				-0.14**
				(0.06)
Prior Week Return		-0.03***	-0.03***	-0.03***
		(0.00)	(0.00)	(0.00)
Abnormal Turnover		-0.04	-0.04	-0.04
		(0.02)	(0.02)	(0.02)
Volatility		1.02	1.07	1.07
		(1.27)	(1.28)	(1.28)
Analyst Revision			0.62***	0.62***
			(0.09)	(0.09)
Earnings Surprise			2.60***	2.60***
			(0.51)	(0.51)
Const.	0.03**	0.01	0.01	0.01
	(0.01)	(0.04)	(0.04)	(0.04)
News Sentiment	No	No	Yes	Yes
Obs.	12633155	12633155	12633155	12633155

Table 3: Regressions of Abnormal Returns on Tweet Sentiment: Stock Subsamples

This table reports the results of regressing abnormal returns over the next week (in percentage points) on tweet sentiment as well as its interaction with three subsamples of stocks: (1) the top 10 most frequently-tweeted stocks, (2) S&P500 stocks not in top 10, and (3) stocks other than those in the first two categories. *Top10*, *S&P500*, and *Others* are dummy variables for these three subsamples. *Sentiment* is defined as the natural log of the ratio of one plus the number of positive tweets to one plus the number of negative tweets, normalized by its standard deviation on days when it is non-zero. Controls are the same as in column 3 of Table 2: they include prior week returns, abnormal turnover, volatility, analyst revisions, earnings surprises, and news sentiment.

	AbnRet
Sentiment	0.20***
	(0.05)
Sentiment × Top10	-0.18
	(0.13)
Sentiment × S&P500	-0.17***
	(0.05)
Top10	0.29***
	(0.08)
S&P500	0.00
	(0.03)
Const.	0.01
	(0.04)
Controls	Yes
Obs.	12633155

Table 4: Tweeting Activity around News Days

This table reports the results of regressing the extensive margin of tweeting activity on whether there was news about the stock using a panel of CRSP stocks from 2008 to 2020. The column headers indicate the dependent variable. *Tweeted Dummy* is a dummy variable for whether the stock was tweeted with a non-neutral sentiment on that day. *Tweet Count* is the log of one plus the total number of positive and negative tweets for that stock on that day. *RecDay* is a dummy variable indicating whether at least one analyst changes her recommendation on that day or the day before. *EarnDay* is a dummy variable indicating whether there was an earnings announcement on that day or the day before. *OtherNewsDay* is a dummy variable indicating whether Ravenpack contains at least one observation about that stock on that day or the day before among the categories listed in Appendix C. *RecSentHigh*, *RecSentMed*, and *RecSentLow* are dummy variables indicating whether the sentiment of the news about analyst recommendation changes was in the top, medium, or bottom terciles of its distribution on that day. The other six independent variables are defined similarly. Standard errors are clustered at the stock level. ***, **, and * represent statistical significance at 1%, 5%, and 10% respectively.

	Tweeted Dummy	Tweet Count	Tweeted Dummy	Tweet Count
Analyst Revision	0.0414***	0.0129***		
	(0.0005)	(0.0002)		
Earnings	0.0176***	0.0058***		
	(0.0006)	(0.0002)		
Other News	0.0018***	0.0006***		
	(0.0001)	(0.0000)		
RecSentHigh			0.0554***	0.0173***
			(0.0013)	(0.0004)
RecSentMed			0.0526***	0.0163***
			(0.0010)	(0.0003)
RecSentLow			0.0522***	0.0163***
			(0.0008)	(0.0003)
EarnSentHigh			0.0117***	0.0038***
			(0.0006)	(0.0002)
EarnSentMed			0.0312***	0.0105***
			(0.0013)	(0.0005)
EarnSentLow			0.0090***	0.0029***
			(0.0005)	(0.0002)
OtherNewsSentHigh			0.0030***	0.0010***
			(0.0002)	(0.0001)
OtherNewsSentMed			0.0010***	0.0003***
			(0.0001)	(0.0000)
OtherNewsSentLow			0.0011***	0.0004***
			(0.0002)	(0.0001)
Average FEs	0.0011	0.0003	0.0012	0.0004
AR Terms	5	5	5	5
Stock FE	Yes	Yes	Yes	Yes
Obs.	12595089	12595089	12595089	12595089

Table 5: Agreement between News and Tweets

This table reports the percentage of tweeted stock/days with non-neutral news based on the sentiment of tweets and news. The first (second) row represents the percentage of negative (positive) stock/days. The last row show the number of non-neutral tweeted stock/days for which there was an event, i.e. a recommendation change, earnings or, other news, either on that day or the day before. Columns below *Analyst Revisions* represents stock/days for which at least one analyst changed her recommendation about the stock either on that day or the day before. Columns below *Earnings* represent stock/days for which there was an earnings announcement for that stock either on that day or the day before. Columns below *Other News* represent stock/days for which there is at least one non-neutral observation in Ravenpack for that stock either on that day or the day before among one of the 24 news categories listed in Appendix C.

	Analyst Revisions		Earnings		Other News	
	Negative	Positive	Negative	Positive	Negative	Positive
Negative	44.84	1.77	17.40	17.16	9.82	14.45
Positive	7.07	46.32	12.67	52.78	29.18	46.56
Obs.	15629		4570		20088	

Table 6: Regressions of Abnormal Returns on Tweet Sentiment: News Subsamples

This table reports the results of regressing abnormal returns over the next week (in percentage points) on tweet sentiment on news days and non-news days. *Sentiment* is defined as the natural log of the ratio of one plus the number of positive tweets to one plus the number of negative tweets, normalized by its standard deviation on days when it is non-zero. *RecDay* is a dummy variable indicating whether at least one analyst revised her recommendation for that stock on that day or the day before. *EarnDay* is a dummy variable indicating whether there was an earnings announcement for that stock on that day or the day before. *OtherNewsDay* is a dummy variable indicating whether there were any news stories other than analyst revisions and earnings announcements for that stock on that day or the day before in the Ravenpack data. Appendix C lists the categories of news included in this variable. *Prior Week Return* is the stock return over the period $[t-4, t]$. *Abnormal Turnover* is the difference between the ratio of the number of shares traded to total shares outstanding at day t and the average of the same ratio over the six-month period ending one month prior to that day. *Volatility* is the standard deviation of daily returns over the month leading to day t. Standard errors are double-clustered at the stock and month levels. ***, **, and * represent statistical significance at 1%, 5%, and 10% respectively.

	(1)	(2)	(3)	(4)
Sentiment	0.13***	0.10**	0.13***	0.14***
	(0.05)	(0.04)	(0.04)	(0.05)
Sentiment × RecDay	-0.05			-0.06
	(0.05)			(0.05)
Sentiment × EarnDay		0.07		0.11
		(0.08)		(0.08)
Sentiment × OtherNewsDay			-0.04	-0.06
			(0.04)	(0.04)
RecSent	0.63***	0.62***	0.62***	0.63***
	(0.08)	(0.09)	(0.09)	(0.09)
EarnSent	2.60***	2.61***	2.60***	2.61***
	(0.51)	(0.51)	(0.51)	(0.51)
OtherNewsSent	0.11***	0.11***	0.11***	0.11***
	(0.02)	(0.02)	(0.02)	(0.02)
RecDay	0.01			0.00
	(0.03)			(0.03)
EarnDay		0.02		0.01
		(0.03)		(0.03)
OtherNewsDay			0.02	0.02
			(0.02)	(0.02)
Prior Week Return	-0.03***	-0.03***	-0.03***	-0.03***
	(0.00)	(0.00)	(0.00)	(0.00)
Abn. Turnover	-0.04	-0.04	-0.04	-0.04
	(0.02)	(0.02)	(0.02)	(0.02)
Volatility	1.08	1.08	1.08	1.08
	(1.28)	(1.28)	(1.28)	(1.28)
Const.	0.01	0.01	0.00	0.00
	(0.04)	(0.04)	(0.04)	(0.04)
Obs.	12633155	12633155	12633155	12633155

Table 7: Regressions of Abnormal Returns on Tweet Sentiment: Deciles of Prior Week Returns

This table reports the results of regressing abnormal returns over the next week (in percentage points) on tweet sentiment for stocks in each decile of prior week returns. Decile breakpoints are calculated using NYSE stocks only. *Sentiment* is defined as the natural log of the ratio of one plus the number of positive tweets to one plus the number of negative tweets. *Abnormal Turnover* is the difference between the ratio of the number of shares traded to total shares outstanding at day t and the average of the same ratio over the six-month period ending one month prior to that day. *Volatility* is the standard deviation of daily returns over the month leading to day t. *Analyst Revision* is the rolling sum over the prior trading week of the change in the stock's analyst recommendations normalized such that a value of 1 represents an upgrade from strong sell to strong buy. *Earnings Surprise* is the rolling sum over the prior trading week of the difference between realized quarterly EPS and IBES analyst consensus forecast normalized to stock price on the announcement day. *News Sentiment* represents 24 control variables, each equal to the rolling sum of the Ravenpack sentiment for the a news category conditional on a novelty and relevance score of at least 75. Each Ravenpack sentiment observation is scaled to the interval $[-1, 1]$. Appendix C includes a list of news categories included in this table. Standard errors are double-clustered at the stock and month levels. ***, **, and * represent statistical significance at 1%, 5%, and 10% respectively.

	Lowest	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	Highest
Sentiment	0.31***	0.02	0.03	0.08	-0.05	-0.10	-0.09	-0.01	0.01	0.17**
	(0.11)	(0.09)	(0.09)	(0.08)	(0.08)	(0.07)	(0.06)	(0.05)	(0.06)	(0.07)
Abnormal Turnover	-0.05*	0.02	0.25***	0.18	0.14***	0.20	0.27	-0.12	-0.30***	-0.04
	(0.03)	(0.04)	(0.08)	(0.12)	(0.04)	(0.36)	(0.22)	(0.13)	(0.09)	(0.03)
Volatility	8.07**	3.27	2.13	-0.25	-1.09	-1.58	-3.63***	-3.65*	-3.25*	-9.08***
	(3.36)	(2.05)	(1.65)	(0.97)	(1.00)	(1.22)	(1.18)	(1.90)	(1.94)	(1.72)
Analyst Revision	0.38**	0.33***	0.39***	0.28***	0.26***	0.25***	0.23***	0.19**	0.33***	0.42***
	(0.17)	(0.10)	(0.11)	(0.08)	(0.07)	(0.08)	(0.08)	(0.08)	(0.08)	(0.09)
Earnings Surprise	2.21***	0.36	2.38**	2.79**	3.25**	3.89***	1.46	3.86***	1.10	1.39
	(0.75)	(0.76)	(1.13)	(1.19)	(1.49)	(1.32)	(1.54)	(1.15)	(0.93)	(0.88)
Const.	0.32**	0.06	0.04	0.10**	0.02	-0.03	-0.00	0.03	-0.01	0.17**
	(0.13)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)	(0.04)	(0.05)	(0.05)	(0.07)
News Sentiment	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	1782994	1309370	1179738	1116421	1085314	1074965	1084159	1122982	1218284	1658627

Table 8: Regressions of Abnormal Returns on Tweet Sentiment: Interactions with Past Returns

This table reports the results of regressing abnormal returns over the next week (in percentage points) on tweet sentiment and its interaction with a dummy for extreme past returns. Column headers indicate the horizon of past returns. *Sentiment* is defined as the natural log of the ratio of one plus the number of positive tweets to one plus the number of negative tweets, normalized by its standard deviation on days when it is non-zero. *Past Return Decile (Highest)* through *Past Return Decile (Lowest)* are dummy variables for deciles of past returns. *Abnormal Turnover* is the difference between the ratio of the number of shares traded to total shares outstanding at day t and the average of the same ratio over the six-month period ending one month prior to that day. *Volatility* is the standard deviation of daily returns over the month leading to day t. *Analyst Revision* is the rolling sum over the prior trading week of the change in the stock's analyst recommendations normalized such that a value of 1 represents an upgrade from strong sell to strong buy. *Earnings Surprise* is the rolling sum over the prior trading week of the difference between realized quarterly EPS and IBES analyst consensus forecast normalized to stock price on the announcement day. *News Sentiment* represents 24 control variables, each equal to the rolling sum of the Ravenpack sentiment for the a news category conditional on a novelty and relevance score of at least 75. Each Ravenpack sentiment observation is scaled to the interval $[-1, 1]$. Appendix C includes a list of news categories included in this table. Standard errors are double-clustered at the stock and month levels. ***, **, and * represent statistical significance at 1%, 5%, and 10% respectively.

	One Day	Two Days	One Week
Sentiment	-0.01	-0.03	-0.02
	(0.03)	(0.03)	(0.03)
Sentiment $\times$ Extreme Past Return	0.31***	0.34***	0.31***
	(0.05)	(0.06)	(0.06)
Past Return Decile (Highest)	-0.36***	-0.35***	-0.29***
	(0.06)	(0.07)	(0.08)
Past Return Decile (9)	-0.06	-0.08*	-0.10**
	(0.04)	(0.04)	(0.05)
Past Return Decile (8)	-0.02	-0.03	-0.05
	(0.04)	(0.04)	(0.04)
Past Return Decile (7)	0.01	-0.01	-0.03
	(0.04)	(0.04)	(0.04)
Past Return Decile (6)	0.03	0.02	0.01
	(0.03)	(0.03)	(0.03)
Past Return Decile (5)	0.04	0.04	0.04
	(0.03)	(0.03)	(0.03)
Past Return Decile (4)	0.07**	0.06*	0.06*
	(0.03)	(0.03)	(0.03)
Past Return Decile (3)	0.07**	0.09***	0.09***
	(0.03)	(0.03)	(0.03)
Past Return Decile (2)	0.14***	0.15***	0.16***
	(0.04)	(0.04)	(0.04)
Past Return Decile (Lowest)	0.49***	0.52***	0.50***
	(0.05)	(0.06)	(0.06)
Abnormal Turnover	-0.05	-0.05	-0.05
	(0.03)	(0.03)	(0.03)
Volatility	-0.39	-0.47	-0.49
	(1.21)	(1.22)	(1.23)
Analyst Revision	0.56***	0.67***	0.58***
	(0.09)	(0.09)	(0.09)
Earnings Surprise	2.34***	2.43***	2.40***
	(0.52)	(0.52)	(0.52)
News Sentiment	Yes	Yes	Yes
Obs.	12633155	12633155	12633155

Table 9: Regressions of Abnormal Returns on Tweet Sentiment: User Subsamples

This table reports the results of regressing abnormal returns over the next week (in percentage points) on tweet sentiment for each adviser type. *SentPlanner* is the aggregate sentiment of financial planners' tweets. Likewise, *SentFundManager* and *SentOtherAdvisers* denote the aggregate sentiment for fund managers and all advisers other than fund managers and financial planners. All sentiment variables are normalized by the standard deviation of sentiment in Table 2 on days when it is non-zero. *Planner*, *FundManager*, and *OtherAdvisers* are dummy variables for stock/days when at least one member of the respective group tweets. *BottomPlanner* (*BottomFundManager*) is a dummy for stock/days when a financial planner (fund manager) other than the top five most frequently tweeting advisers tweets. Controls are the same as in column 3 of Table 2. Standard errors are double-clustered at the stock and day levels. ***, **, and * represent statistical significance at 1%, 5%, and 10% respectively.

	(1)	(2)
SentPlanner	0.12*	0.10
	(0.07)	(0.07)
SentFundManager	0.06	0.11*
	(0.08)	(0.06)
SentOtherAdvisers	0.09	0.09
	(0.38)	(0.39)
SentPlanner × BottomPlanner		0.30
		(0.38)
SentFundManager × BottomFundManager		-0.28
		(0.25)
Planner	0.11	0.12
	(0.08)	(0.08)
BottomPlanner		-0.08
		(0.60)
FundManager	-0.08	-0.07
	(0.10)	(0.07)
BottomFundManager		-0.06
		(0.35)
OtherAdvisers	0.02	0.02
	(0.42)	(0.42)
Const.	0.01	0.01
	(0.04)	(0.04)
Controls	Yes	Yes
Obs.	12633155	12633155

Table 10: Regressions of Retail Order Imbalance on Tweet Sentiment

This table reports the results of the regression

$$RetailOIB_{i,t+1,t+5} = \beta Sentiment_{i,t} + \gamma X_{i,t} + \eta_i + \xi_t + \epsilon_{i,t+1,t+5},$$

on a panel of CRSP stocks. The sample spans from January 1, 2010 to December 31, 2020 in columns 1-4 and to December 31, 2015 in column 5. *RetailOIB* is defined as the standardized ratio of the difference between number of shares bought and sold by retail investors to their sum. *Sentiment* is defined as the natural log of the ratio of one plus the number of positive tweets to one plus the number of negative tweets. *Positive* is defined as the maximum of Sentiment and zero. *Negative* is the maximum of -1 times Sentiment and zero. Sentiment, Positive, and Negative are normalized by the standard deviation of Sentiment on days when it is non-zero. Control variables are prior week returns, abnormal turnover, volatility, analyst revisions, earnings surprises, news sentiment, and retail order imbalance over the interval $[t-4, t]$. Standard errors are double-clustered at the stock and day levels. ***, **, and * represent statistical significance at 1%, 5%, and 10% respectively.

	(1)	(2)	(3)	(4)	(5)
Sentiment	0.023***	0.024***	0.026***		0.023***
	(0.003)	(0.003)	(0.003)		(0.004)
Positive				0.029***	
				(0.004)	
Negative				-0.015***	
				(0.006)	
Average FEs	-0.0001	-0.0001	0.0065	0.0065	0.0050
Stock FE	Yes	Yes	Yes	Yes	Yes
Day FE	No	Yes	Yes	Yes	Yes
Controls	No	No	Yes	Yes	Yes
Obs.	9026641	9026641	9026641	9026641	5159123

A. Twitter Data

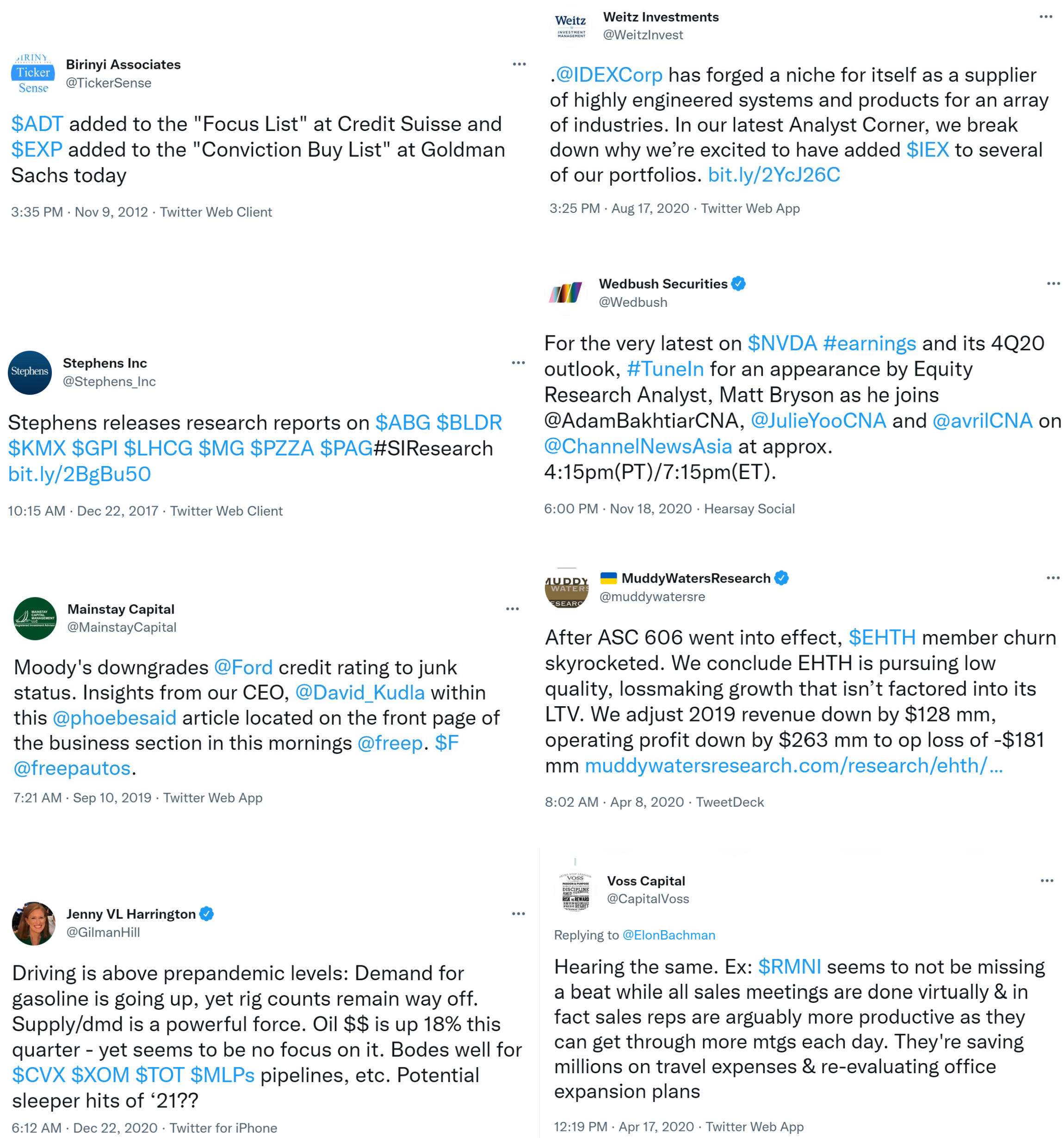


Figure A.1: Examples of RIA Tweets

Table A.1: Summary Statistics of Advisers' Twitter Accounts

This table describes Twitter profiles of RIA firms as of the time of data collection in May 2021, regardless of whether they tweet about stocks. *Followers* is the number of followers. *Years Active* is the number of years the account has been active. *Tweets* is the users' total number of tweets. *Stock Tweets* is the number of tweets that mention at least one CRSP stock.

	N	Mean	SD	P1	P10	P25	P50	P75	P90	P99
Followers	697	16810.78	71993.29	14	62	178	773	4352	26235	387394
Years Active	697	8.07	2.93	0.90	3.76	6.23	8.38	10.26	11.82	12.53
All Tweets	697	3911.38	7765.50	24	187	503	1400	4050	9965	37555
Cashtags	697	243.16	1389.97	1	1	2	5	29	238	6836

Table A.2: Summary Statistics of Stock Tweets

This table describes tweets mentioning at least one CRSP stock with share code 10, 11, or 12 and exchange code 1, 2, or 3. Neutral tweets are included. All variables are reported as of the time of data collection in May 2021.

	N	Mean	SD	P1	P10	P25	P50	P75	P90	P99
Likes	99798	13.71	129.03	0	0	0	1	4	15	204
Quotes	99798	0.33	3.15	0	0	0	0	0	1	6
Replies	99798	1.07	6.91	0	0	0	0	1	2	17
Retweets	99798	5.13	50.98	0	0	0	0	2	8	79
Cashtags	99798	1.70	2.32	1	1	1	1	1	3	13

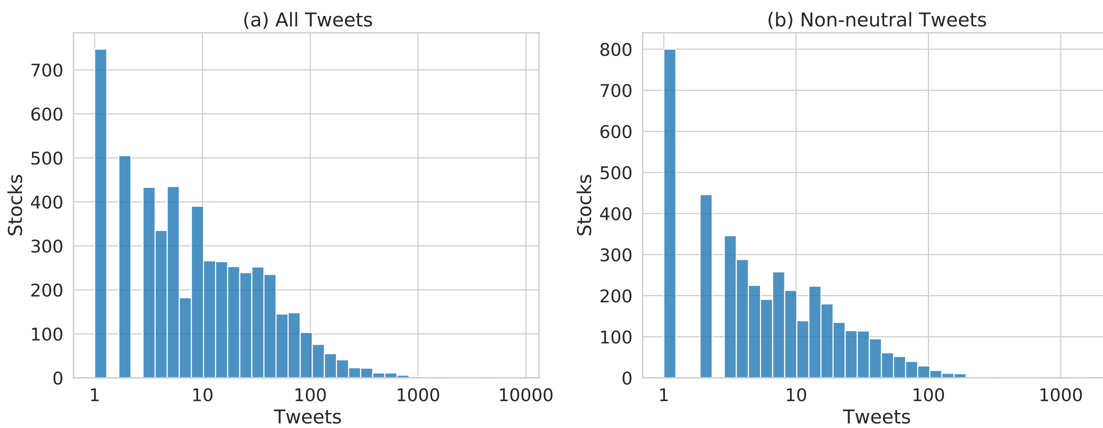


Figure A.2: The Distribution of Tweets among Stocks

This figure shows a histogram of the number of times each stock is tweeted by investment advisers.

B. Description of NLP Models

Most authors use bag-of-words algorithms to assign sentiment to texts. These methods take words as independent draws from some distribution of sentiment and aggregate each word's sentiment to obtain that of the sentence. Although very popular in the literature, bag-of-words methods have several drawbacks. First, the sequence of words in a sentence matters for its overall sentiment. For example, "Buy IBM now and sell it later" is a positive sentence while "Sell IBM now and buy it later", though using the same words, is negative. Second, the sentiment of a sentence is often not the aggregate sentiment of its words. A sentence such as "You should not sell IBM" has a positive sentiment despite containing two negative words.

Furthermore, bag-of-words methods cannot handle multiple stocks mentioned in the same sentence. For example, "upgrade IBM and downgrade INTC" has a positive sentiment for IBM and a negative sentiment for INTC. At the same time, bag-of-words methods can only attribute one sentiment to the sentence. As a result, most papers drop tweets with multiple cashtags from their samples, leading to considerable data losses. I estimate that 53.2% of the data comes from tweets with multiple cashtags. Moreover, the information content of tweets with a single cashtag is different from those with multiple cashtags. For example, a casual reading of tweets shows that technical analysts often use one cashtag per tweet, while analyst recommendations about multiple stocks are often bundled together in one tweet. Therefore, dropping tweets with multiple cashtags can lead to biased results. Loughran and McDonald (2016) offer a more detailed discussion on bag-of-words methods.

To address such concerns, I used two state-of-the-art deep learning algorithms to assign sentiment to tweets; namely, LSTM and BERT models. Both algorithms are used to analyze financial text (Azimi and Agrawal (2021), De la Pera (2021), Dangl and Salbrechter (2021)). LSTM models are regular neural networks preceded by a layer of recurrent neurons to remember the order of words. BERT models are state-of-the-art structures used by Google for processing search queries. Both models are sensitive to the sequence of words, which enables me to avoid the disadvantages of bag-of-word methods. The next two sections describe the main ideas behind the LSTM and BERT models.

B.1. LSTM Models

The sequence and context of words are important for understanding human languages. Therefore, any algorithm made for the purpose of understanding human languages should be able to interpret each word with relation to its neighboring words. In deep learning, Recurrent Neural Networks (RNNs), whose structure is shown in Figure B.3, are used for such applications. An RNN is simply a sequence of cells, each of which takes as input an embedded word and a state variable. The state variable acts as the context for the word being processed. Each cell processes the input word using the state variable and weights learned during training and generates an output. Furthermore, it also modifies the state variable and passes it to the next cell. Therefore, we can sequentially feed the words of a sentence to a chain of RNN cells. The RNN cells will then generate a processed version of the sentence, which can be fed into another neural network adapted to the specific task in hand.

Even though RNNs can theoretically remember the context in arbitrarily long sequences of words, in practice they suffer from the vanishing gradient problem. In other words, the gradient of the error function becomes vanishingly small during training, effectively stopping the algorithm from making progress. To solve the vanishing gradient problem, a specific structure of RNNs called Long Short-Term Memory (LSTM) networks are used. An LSTM cell passes the unmodified input state to the next cell in the chain which helps the network "remember" the context in very long sequences. Because the context of a word can come either before or after its position in the sentence, NLP models sometimes feature two LSTM blocks; one reading the sentence from left to right and the other one from right to left. Such a structure is called bidirectional LSTM. Furthermore, an attention layer is placed after the LSTM layer that weighs the words in a sentence based on a pattern of weights learnt during the training process. Figure B.4 shows the structure of a typical LSTM model used in NLP.

B.2. BERT Models

The field of natural language processing has progressed rapidly over the past few years. At the core of this progress were two central ideas. The first idea was a superior structure. In 2017, Vaswani et al. (2017) showed that an attention-based structure can achieve a higher accuracy in English-to-German translation than the best models of its time. They named this structure a "Transformer". Transformers rapidly gained popularity for NLP applications not only due to their superior performance, but also because their parallel structure meant it took only a fraction of time to train them. Hence, they could be trained on more data at the same cost.

The second idea was pretraining. Some machine learning applications, like NLP, have an inherent structure that does not change across tasks. Therefore, one can

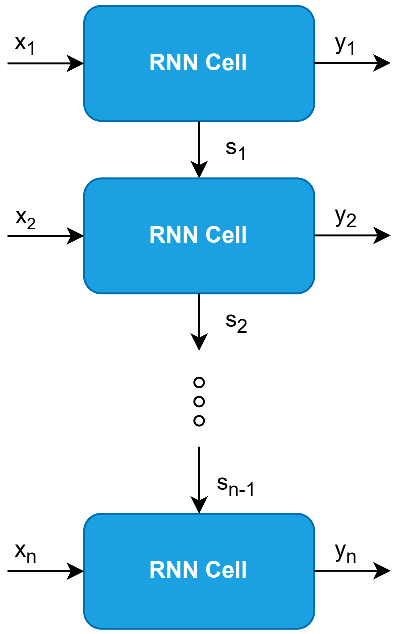


Figure B.3: A Recurrent Neural Network

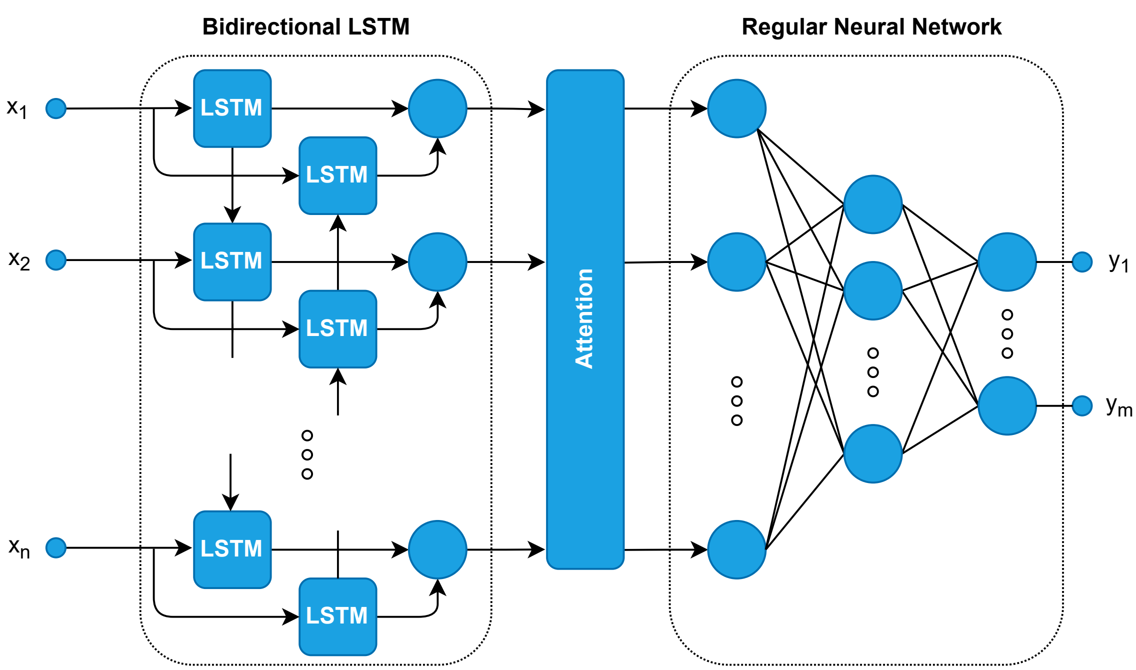


Figure B.4: The Typical Structure of an LSTM Model Used in NLP train an algorithm on one task and make minor modifications to use the model for a separate task. For example, once an NLP model is trained on parts-of-speech tagging, we may use the learned weights to initialize the training of the model for sentiment analysis. The first step is called pretraining while the second fine-tuning. Pretraining often improves the model's performance while using fewer data.

Because it takes much less time to train a transformer, we can pretrain them on large corpora such as Wikipedia, which allows transformers to achieve even better performance metrics. In 2018, Devlin et al. (2018) created pretrained NLP models based on transformers, which they called BERT. They use two pretraining methods. First, they randomly maske words in their corpus and let the transformer predict the masked word in the sentence. Second, they pick random pairs of sentences from the corpus. For some of these pairs, the second sentence immediately follows the first one. For each pair, the model has to predict whether the second sentence follows the first. Devlin et al. (2018) mention that once pretarining is finished, the fine tuning can be as simple as adding a single layer to the BERT model. They fine-tune the BERT models for several benchmark NLP tasks and show that they outperform state-of-the-art models in each case. In this paper, I fine-tuned small BERT models to create my sentiment analysis algorithms.

Figure B.5 illustrates the basic structure of a BERT model. Words enter the embedding layer at the left side of the structure. The embedding layer converts words and their positions into a vector of real numbers whose dimensions are the same as the model. An attention mechanism is then applied to the output of the encoding layer whose weights are learned during the training phase. The output of

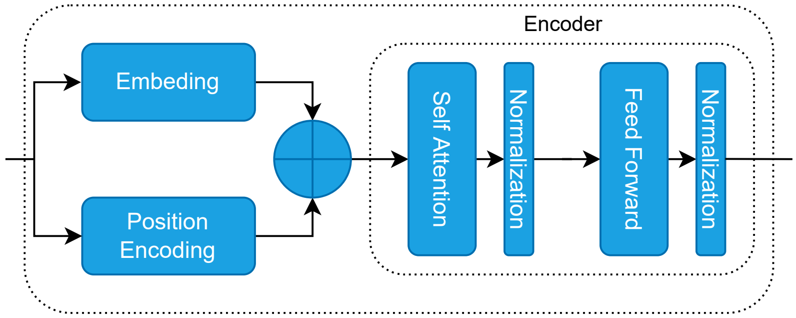


Figure B.5: The Basic Structure of a BERT Model

the attention layer is then normalized and passed on to a regular neural network, whose outputs are again normalized and passed as the output of the entire model. In practice, BERT models use several stages of this structure. Moreover, the attention mechanism is often broken into several attention heads to accelerate the training. Readers who would like to learn more about BERT models can refer to Rothman (2021).

B.3. Applying NLP Models to Tweets

To assign sentiment to tweets with multiple cashtags, I generate a separate observation for each cashtag by coding that it differently from others. For example, for a tweet such as "upgrade \$IBM, downgrade \$INTC", to measure the sentiment with respect to IBM, I change the text of the tweet to "upgrade *[main]*, downgrade *[other]*". The placeholder *[main]* is a special word replacing the stock for which I want to measure the sentiment. Likewise, *[other]* replaces all other stocks in the tweet. Hence the same tweet converts to "upgrade *[other]*, downgrade *[main]*" when I want to measure the sentiment for INTC. Therefore, each tweet generates as many observations as the number of stocks it mentions. After training the models, I confirmed the validity of my approach in three different ways. First, I manually verified that the algorithms assigned different sentiments to at least some of the tweets such as above. Second, I trained my models after removing irrelevant cashtags from the tweets. Third, I trained my algorithms with only single-cashtag tweets. In the last two exercises, I get similar accuracies as in my main approach.[3] Next, I manually labeled 6148 tweets as positive, neutral, or negative and broke them into a training set ($n = 4148$), a validation set ($n = 1000$), and a test set ($n = 1000$).

After experimenting with the details of the models, I chose the following networks based on total accuracy over the validation set:[4]

1. LSTM: An LSTM layer with 32 neurons, followed by two dense layers with 16 and 8 neurons,
2. BERT256: Small BERT with parameters (L=2, H=256, A=4),
3. BERT512: Small BERT with parameters (L=2, H=512, A=8).

Figure B.6 shows the two ways I converted the networks above to sentiment analysis algorithms. First, I added a three-neuron dense layer to the network to develop a three-class classifier. Second, I added a single neuron with linear activation to each network and build a regression model. Let x denote the output of the regression model. I converted x to a sentiment label by comparing it to a threshold

[3] When training models on single-cashtag tweets, I had to drop the rest of the tweets. As a result, the training set was smaller.

[4] One can download the BERT networks used in this paper as part of the library *tensorflow_hub* for Python.

T according to the following function:

$$\text{Sent}(x) = \begin{cases} \text{positive} & T \leq x, \\ \text{neutral} & -T < x < T, \\ \text{negative} & x \leq -T. \end{cases} \tag{B.1}$$

Notice that increasing the threshold implies the regression model will tag more tweets as neutral. In that sense, a higher threshold makes the model more conservative. From each network, I created three sentiment analysis algorithms: a classifier and two regression models with thresholds of 0.5 and 0.8.

Aggregating predictions of several models (ensemble modeling) can increase accuracy compared to each model. The intuition behind this combination is that each model's prediction is a noisy signal of the true sentiment. Therefore, the aggregate signal will be more informative than each signal alone. I find the most frequent label among all algorithms for each observation and call it the consensus algorithm.

I use three metrics for each model's performance:

1. Accuracy: the percentage of all tweets correctly classified,
2. Precision: the percentage of correctly classified tweets given the predicted label.
3. Recall: the percentage of tweets correctly classified given the true label.

Table B.3 reports these three metrics for each of the ten models. In the test set, 60.2% of all tweets are neutral, 26.2% are positive, and 13.6% are negative. Overall, the consensus model has the highest accuracy (83.1%), in line with the intuition behind ensemble modeling. It also achieves a precision of more than 80% in all sentiment categories. Relatively low recall figures for the negative class imply that

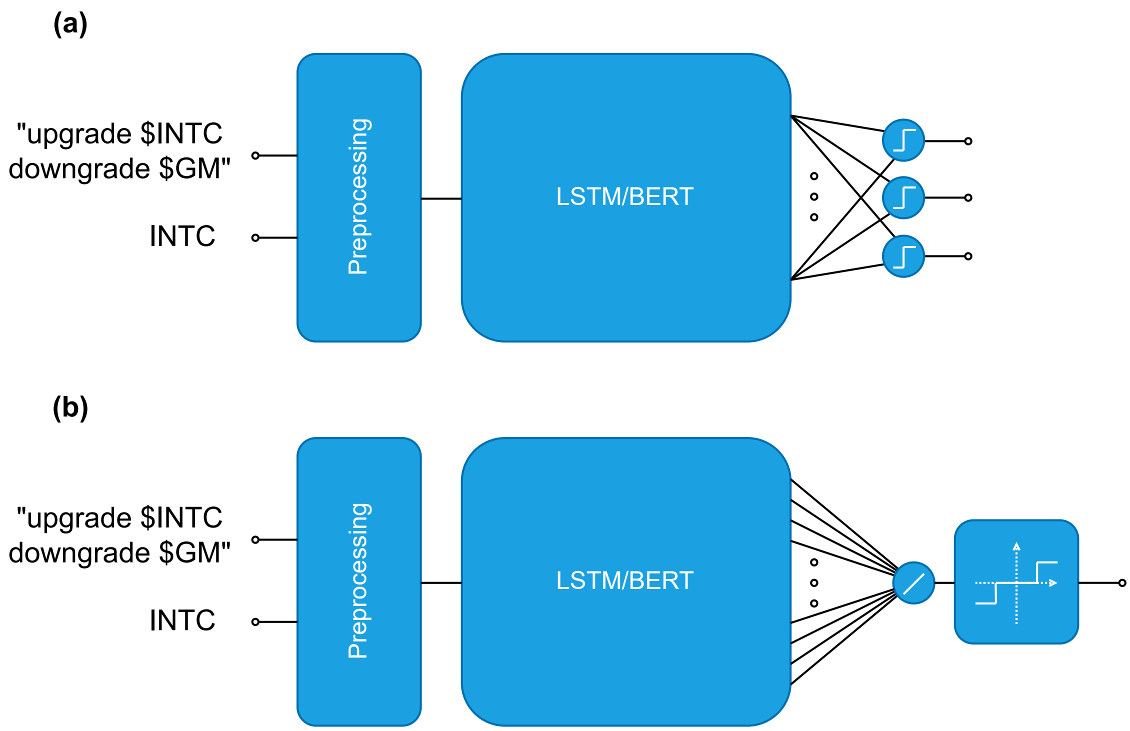


Figure B.6: Sentiment Analysis Using NLP Models

I build sentiment analysis algorithms by adding extra layers to the end of a general purpose LSTM or BERT model. (a) The structure of a classification model (b) The structure of a regression model.

all models struggle to tag negative tweets. This result is not surprising given the low number of negative tweets in the training sample.

Two models with an identical accuracy may misclassify the same subset of tweets or two disjoint subsets. Therefore, it is natural to ask how much the models agree with one another. Table B.4 presents the percentage of all tweets on which each pair of algorithms agree. The lowest agreement (69.8%) is between the LSTM classifier and the BERT256 regression model with a threshold of 0.5. The highest agreement (91.4%) is between the two LSTM regression models. We can also measure the correlation between the outcomes. Table B.5 presents the correlation matrix. The lowest correlation (0.41) is between the LSTM classifier and the BERT256 regression models classifiers, as well as the BERT512 classifier model. The highest correlation (0.88) is between the two LSTM regression models. Both tables indicate that even though the models agree on the majority of tweets, there is enough disagreement to justify aggregating the labels.

Table B.3: The Performance of Sentiment Analysis Algorithms

All models are trained on the same random sample of 5148 tweets (4148 training, 1000 validation) and tested on another random sample of 1000 tweets. *Accuracy* is defined as the fraction of tweets correctly classified. *Precision* is the ratio TP/(TP+FP). *Recall* is the ratio TP/(TP+FN). TP, FP, and FN represent true positive, false positive, and false negative, where positives and negatives should not be confused with the sentiment labels. All performance measures are reported in percentage points. The highest number in each column is in bold. The last row indicates the number of total, true negative, true neutral, and true positive observations in the test set.

	Total	Negative		Neutral		Positive	
	Accuracy	Precision	Recall	Precision	Recall	Precision	Recall
LSTM Classifier	75.0%	81.3%	44.9%	**86.1%**	80.1%	56.7%	**79.0%**
LSTM Regression, Threshold = 0.5	77.2%	67.3%	52.9%	83.7%	82.1%	68.0%	78.6%
LSTM Regression, Threshold = 0.8	80.3%	85.0%	50.0%	81.9%	90.2%	74.7%	73.3%
BERT256 Classifier	81.8%	**95.5%**	46.3%	81.8%	**94.2%**	78.0%	71.8%
BERT256 Regression, Threshold = 0.5	76.2%	73.8%	58.1%	81.9%	82.7%	64.9%	70.6%
BERT256 Regression, Threshold = 0.8	78.6%	92.3%	44.1%	77.7%	92.9%	77.3%	63.7%
BERT512 Classifier	82.2%	81.8%	**59.6%**	86.2%	88.9%	73.6%	78.6%
BERT512 Regression, Threshold = 0.5	77.8%	68.7%	58.1%	82.4%	84.7%	71.1%	72.1%
BERT512 Regression, Threshold = 0.8	80.2%	74.3%	57.4%	79.2%	93.0%	**87.2%**	62.6%
Consensus	**83.1%**	91.2%	53.7%	82.4%	94.0%	82.4%	73.3%
Obs.	1000	136		602		262	

Table B.4: The Agreement Matrix among All Sentiment Analysis Models

In this matrix, each element represents the percentage of tweets to which the models in the row and column assign the same label. All tweets (N=169486) have been used in calculating this matrix. The algorithms are listed as follows: (1) LSTM classifier (2) LSTM regression model with threshold=0.5 (3) LSTM regression model with threshold=0.8 (4) BERT256 classifier (5) BERT256 regression model with threshold=0.5 (6) BERT256 regression model with threshold=0.8 (7) BERT512 classifier (8) BERT512 regression model with threshold=0.5 (9) BERT512 regression model with threshold=0.8 (10) Consensus model.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1) LSTM	100.0	72.7	75.0	74.7	69.8	69.8	72.9	70.1	70.6	78.6
(2) LSTM (0.5)		100.0	91.4	75.0	73.0	72.2	75.2	73.5	73.9	83.6
(3) LSTM (0.8)			100.0	80.0	76.6	78.7	77.3	77.2	79.9	89.5
(4) BERT256				100.0	75.9	79.3	83.4	75.5	79.3	87.1
(5) BERT256 (0.5)					100.0	85.5	73.8	75.5	77.8	83.8
(6) BERT256 (0.8)						100.0	72.9	76.5	82.7	86.3
(7) BERT512							100.0	73.4	74.4	83.0
(8) BERT512 (0.5)								100.0	88.6	83.7
(9) BERT512 (0.8)									100.0	87.0
(10) Consensus										100.0

Table B.5: The Correlation Matrix among All Sentiment Analysis Models

To calculate correlations, I code positive, neutral, and negative tweets as +1, 0, and -1 respectively. All tweets (N=169486) have been used in calculating this matrix. The algorithms are listed as follows: (1) LSTM classifier (2) LSTM regression model with threshold=0.5 (3) LSTM regression model with threshold=0.8 (4) BERT256 classifier (5) BERT256 regression model with threshold=0.5 (6) BERT256 regression model with threshold=0.8 (7) BERT512 classifier (8) BERT512 regression model with threshold=0.5 (9) BERT512 regression model with threshold=0.8 (10) Consensus model.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1) LSTM	1.00	0.45	0.50	0.47	0.41	0.41	0.41	0.43	0.45	0.58
(2) LSTM (0.5)		1.00	0.88	0.51	0.58	0.54	0.59	0.58	0.59	0.75
(3) LSTM (0.8)			1.00	0.57	0.60	0.58	0.61	0.62	0.63	0.81
(4) BERT256				1.00	0.51	0.55	0.61	0.53	0.57	0.67
(5) BERT256 (0.5)					1.00	0.76	0.57	0.60	0.62	0.71
(6) BERT256 (0.8)						1.00	0.55	0.57	0.62	0.71
(7) BERT512							1.00	0.55	0.57	0.70
(8) BERT512 (0.5)								1.00	0.83	0.71
(9) BERT512 (0.8)									1.00	0.74
(10) Consensus										1.00

C. The Data Cleaning Process and Variable Definitions

C.1. CRSP

From daily CRSP files, I keep US and Foreign common stocks, share codes 10, 11, and 12, with exchange codes 1, 2, or 3. To calculate abnormal returns, I divide stocks into 5 groups based on size quintiles of NYSE stocks. Within each group, I divide stocks again based on book-to-market quintiles of NYSE stocks and repeat the process for momentum as well. The result is 125 banchmark portfolios. The forward abnormal return of every stock is the forward return of that stock minus the value weighted forward return of its benchamrk portfolio. I calculate abnormal returns for each stock at horizons up to three months. In addition to abnormal returns, I calculate the following variables:

- Momentum: the return over the prior the period $[t-252, t-21]$.
- Book to market: book equity calcualted as in Davis et al. (2000), divided by market value of equity from CRSP.
- Abnormal turnover: the ratio of total trading volume to shares outstanding minus the average of the same ratio for the same stock over the period $[t-126, t-21]$.
- Volatility: standard deviation of daily returns over the prior 21 trading days.

C.2. TAQ

I calculate intraday prices at 5-minute intervals using the code from Holden and Jacobsen (2014) provided on Craig Holden's website. I drop after-hour and opening prices from the data. Following Boehmer et al. (2021), I identify retail buys/sells as trades whose execution price in dollars can be placed in an interval $(p/100 - 0.005, p/100)/(p/100, p/100 + 0.005)$ where p is an integer. I calculate the total number of shares bought and sold by retail traders at half-hour intervals during each trading day and also at the daily level for each ticker in TAQ. Retail order imbalance for any interval Δt (half-hour, one day, or one week) is defined as

$$RetailOIB_{i,\Delta t} = \frac{RetailBuys_{i,\Delta t} - RetailSells_{i,\Delta t}}{RetailBuys_{i,\Delta t} + RetailSells_{i,\Delta t}}.$$

C.3. Ravenpack

For each stock and day, I keep all news stories with relevance and novelty scores of at least 75 from Dow-Jones equity files. I convert the time stamp in the Ravenpack files to the US/Eastern time zone and assign each news to the first trading day that closes after it. I add the sentiment of news for each news category every day and merge the data with CRSP using the mapping files provided on WRDS. The following categories are included in my main results: mergers and acquisitions, assets, bankruptcy, corporate responsibility, credit, credit ratings, crime, dividends, equity actions, exploration, indexes, industrial accidents, insider trading, investor relations, labor issues, legal, marketing, partnerships, price targets, products services, regulatory, revenues, security, and transportation.

C.4. IBES

I use IBES for two purposes. To calculate analyst recommendation changes, I start from the recommendation detail file and drop observations for which one of the following variables are missing: analyst code, recommendation, announcement date, and announcement time. I drop observations with a recommendation value of zero and calculate the change in recommendation for each analyst/stock pair. Because the value of the recommendation can be between 1 and 5, its change is between -4 and 4. I divide the change in recommendation by 4 to make sure my measure of analyst recommendations is between -1 and 1. To each observation, I assign as date the first trading day closing after the time of the recommendation. I use the linking table provided by WRDS to get the corresponding permno for each IBES ticker. I proceed to merge the data with CRSP on permno and date.

To calculate earning surprises, I read in IBES consensus forecast and actual EPS files for both US and international stocks. For each stock and quarter, I keep the last forecast for quarterly EPS in the period $[t-90, t-1]$, where t is the day of the announcement, and in the same currency as the earnings. I fetch cumulative adjustment factor for prices from CRSP daily files. To correct for any stock splits between the forecast and the earnings announcement, I multiply the forecast by the adjustment factor at the time of the earnings announcement and divide by the factor at the time of the forecast. Finally, I set the earnings surprise equal to the difference between the realized EPS and the mean of EPS forecasts normalized by the stock price at the close of the earnings announcement day. The merge between this file and CRSP is similar to the recommendation files.

C.5. Form ADV

The data from form ADV submissions is divided into several parts, all of which can be downloaded from the SEC website. For the purpose of this paper, I work with part 1 which is available at `https://www.sec.gov/foia/docs/adv/form-adv-complete-ria.zip`. Websites and social media handles are in schedule D item

1I. I obtain a full list of Twitter handles and manually check them for misspellings, repetitions, and other obvious human errors. After cleaning the handles, I search for each user handle and download its profile as well as all the tweets from its timeline. To match user profiles with advisers, I create a linking table containing user profiles and CRD numbers of the advisers mentioning them. The data on the types of services advisers offer are in item 5G. The indicator for whether the adviser advises a private fund comes from item 7.

D. Other Results

Table D.1: Regressions of Abnormal Returns on Tweet Sentiment: Only Tweeted Stock/Days

This table reports the results of regressing abnormal returns over the next week (in percentage points) on sentiment for stock/days tweeted with a non-neutral sentiment. *Sentiment* is defined as the natural log of the ratio of one plus the number of positive tweets to one plus the number of negative tweets. *Positive* is defined as the maximum of sentiment and zero. *Negative* is the maximum of -1 times sentiment and zero. These three variables are normalized by the standard deviation of Sentiment on days when it is non-zero. *Prior Week Return* is the stock return over the period $[t-4,t]$. *Abnormal Turnover* is the difference between the ratio of the number of shares traded to total shares outstanding at day t and the average of the same ratio over the six-month period ending one month prior to that day. *Volatility* is the standard deviation of daily returns over the month leading to day t. *Analyst Revision* is the rolling sum over the prior trading week of the change in the stock's analyst recommendations normalized such that a value of 1 represents an upgrade from strong sell to strong buy. *Earnings Surprise* is the rolling sum over the prior trading week of the difference between realized quarterly EPS and IBES analyst consensus forecast normalized to stock price on the announcement day. *News Sentiment* represents 24 control variables, each equal to the rolling sum of the Ravenpack sentiment for the a news category conditional on a novelty and relevance score of at least 75. Each Ravenpack sentiment observation is scaled to the interval $[-1,1]$. Appendix C includes a list of news categories included in this table. Standard errors are double-clustered at the stock and month levels. ***, **, and * represent statistical significance at 1%, 5%, and 10% respectively.

	(1)	(2)	(3)	(4)
Sentiment	0.10***	0.11**	0.11**	
	(0.03)	(0.04)	(0.05)	
Positive				0.19
				(0.17)
Negative				-0.02
				(0.18)
Prior Week Return		-0.00	-0.00	-0.00
		(0.01)	(0.01)	(0.01)
Abnormal Turnover		-1.43**	-1.45**	-1.45**
		(0.62)	(0.61)	(0.61)
Volatility		-1.31	-2.67	-2.67
		(5.66)	(5.68)	(5.68)
Analyst Revision			-0.06	-0.05
			(0.12)	(0.12)
Earnings Surprise			4.91	4.91
			(3.22)	(3.22)
Const.	-0.06	-0.00	0.01	-0.09
	(0.04)	(0.12)	(0.12)	(0.22)
News Sentiment	No	No	Yes	Yes
Obs.	46241	46241	46241	46241

Table D.2: Regressions of Fama-French (2015) CARs on Tweet Sentiment

This table reports the results of regressing cumulative abnormal returns over the next week (in percentage points) on sentiment for stock/days tweeted with a non-neutral sentiment. Betas on day t are estimated by regressing daily excess returns on factors (market, size, value, investment, and profitability) from Kenneth French's library over the period $[t-273, t-21]$. *Sentiment* is defined as the natural log of the ratio of one plus the number of positive tweets to one plus the number of negative tweets. *Positive* is defined as the maximum of sentiment and zero. *Negative* is the maximum of -1 times sentiment and zero. These three variables are normalized by the standard deviation of Sentiment on days when it is non-zero. *Prior Week Return* is the stock return over the period $[t-4, t]$. *Abnormal Turnover* is the difference between the ratio of the number of shares traded to total shares outstanding at day t and the average of the same ratio over the six-month period ending one month prior to that day. *Volatility* is the standard deviation of daily returns over the month leading to day t. *Analyst Revision* is the rolling sum over the prior trading week of the change in the stock's analyst recommendations normalized such that a value of 1 represents an upgrade from strong sell to strong buy. *Earnings Surprise* is the rolling sum over the prior trading week of the difference between realized quarterly EPS and IBES analyst consensus forecast normalized to stock price on the announcement day. *News Sentiment* represents 24 control variables, each equal to the rolling sum of the Ravenpack sentiment for the a news category conditional on a novelty and relevance score of at least 75. Each Ravenpack sentiment observation is scaled to the interval $[-1, 1]$. Appendix C includes a list of news categories included in this table. Standard errors are double-clustered at the stock and month levels. ***, **, and * represent statistical significance at 1%, 5%, and 10% respectively.

	(1)	(2)	(3)	(4)
Sentiment	0.01	0.15***	0.09**	
	(0.03)	(0.04)	(0.04)	
Positive				0.07
				(0.05)
Negative				-0.18**
				(0.07)
Prior Week Return		-0.03***	-0.03***	-0.03***
		(0.01)	(0.00)	(0.00)
Abnormal Turnover		-0.04*	-0.04*	-0.04*
		(0.02)	(0.02)	(0.02)
Volatility		6.90***	6.90***	6.90***
		(2.62)	(2.62)	(2.62)
Analyst Revision			0.64***	0.64***
			(0.10)	(0.10)
Earnings Surprise			2.23***	2.23***
			(0.51)	(0.51)
Const.	0.11***	-0.10*	-0.10*	-0.10*
	(0.04)	(0.06)	(0.06)	(0.06)
News Sentiment	No	No	Yes	Yes
Obs.	12811719	12811719	12811719	12811719

Table D.3: Regressions of Abnormal Returns on Tweet Sentiment: All Algorithms

This table reports the results of regressing abnormal returns over the next week (in percentage points) on sentiment. *Sentiment* is defined as the natural log of the ratio of one plus the number of positive tweets to one plus the number of negative tweets, normalized to have a standard deviation of one on non-neutral stock/days. The algorithms used in each column are: (1) LSTM classifier (2) LSTM regression model with threshold=0.5 (3) LSTM regression model with threshold=0.8 (4) BERT256 classifier (5) BERT256 regression model with threshold=0.5 (6) BERT256 regression model with threshold=0.8 (7) BERT512 classifier (8) BERT512 regression model with threshold=0.5 (9) BERT512 regression model with threshold=0.8 (10) Consensus model. Similar to Table 2, control variables are prior week returns, abnormal turnover, volatility, analyst revision, earnings surprise, and news sentiment. Standard errors are double-clustered at the stock and month levels. ***, **, and * represent statistical significance at 1%, 5%, and 10% respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Sentiment	0.07**	0.12***	0.13***	0.09***	0.11***	0.12***	0.12***	0.12***	0.14***	0.12***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.04)	(0.04)
Const.	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	12633155	12633155	12633155	12633155	12633155	12633155	12633155	12633155	12633155	12633155

Table D.4: Regressions of Abnormal Returns on Tweet Sentiment: Alternative Horizons

This table reports the results of regressing abnormal returns over the next week (in percentage points) on sentiment, for horizons from one day to three months. *Sentiment* is defined as the natural log of the ratio of one plus the number of positive tweets to one plus the number of negative tweets, normalized to have a standard deviation of one on non-neutral stock/days. Column headers are the horizon of abnormal returns in trading days. Similar to Table 2, control variables are prior week returns, abnormal turnover, volatility, analyst revision, earnings surprise, and news sentiment. Standard errors are double-clustered at the stock and month levels. ***, **, and * represent statistical significance at 1%, 5%, and 10% respectively.

	1	3	5	10	21	42	63
Sentiment	0.06***	0.10***	0.12***	0.14***	0.17*	0.45***	0.66**
	(0.02)	(0.03)	(0.04)	(0.05)	(0.09)	(0.17)	(0.26)
Const.	-0.01	-0.01	0.01	0.09**	0.17**	0.37***	0.50***
	(0.01)	(0.03)	(0.04)	(0.04)	(0.07)	(0.11)	(0.13)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	12633155	12633155	12633155	12633155	12633155	12633155	12633155

Table D.5: Regressions of Abnormal Returns on Tweet Sentiment

This table reports the results of the regression

$$AbnRet_{i,t+1,t+5} = \alpha + \beta Sentiment_{i,t} + \mathbb{1}\{Tweeted\} + \gamma X_{i,t} + \epsilon_{i,t+1,t+5},$$

on a panel of CRSP stocks spanning from January 1, 2008 to December 31, 2020. *AbnRet* are calculated using the method of Daniel et al. (1997) and in percentage points. *Sentiment* is defined as the natural log of the ratio of one plus the number of positive tweets to one plus the number of negative tweets. *Positive* is defined as the maximum of sentiment and zero. *Negative* is the maximum of -1 times sentiment and zero. These three variables are normalized by the standard deviation of Sentiment on days when it is non-zero. *Prior Week Return* is the stock return over the period $[t-4, t]$. *Abnormal Turnover* is the difference between the ratio of the number of shares traded to total shares outstanding at day t and the average of the same ratio over the six-month period ending one month prior to that day. *Volatility* is the standard deviation of daily returns over the month leading to day t. *Analyst Revision* is the rolling sum over the prior two days of the analyst revisions normalized such that a value of 1 represents an upgrade from strong sell to strong buy. *Earnings Surprise* is the sum over the prior two days of the difference between realized quarterly EPS and IBES analyst consensus forecast normalized to stock price on the announcement day. *Tweeted Day* is a dummy variable indicating whether the stock was tweeted on that day. *News Sentiment* represents 24 control variables, each equal to the rolling sum over the prior two days of the Ravenpack sentiment for the a news category conditional on a novelty and relevance score of at least 75. Each Ravenpack sentiment observation is scaled to the interval $[-1, 1]$. Appendix C includes a list of news categories included in this table. Standard errors are double-clustered at the stock and month levels. ***, **, and * represent statistical significance at 1%, 5%, and 10% respectively.

	(1)	(2)	(3)	(4)
Sentiment	0.10***	0.21***	0.13***	
	(0.03)	(0.04)	(0.04)	
Positive				0.23
				(0.17)
Negative				-0.02
				(0.18)
Prior Week Return		-0.03***	-0.03***	-0.03***
		(0.00)	(0.00)	(0.00)
Abnormal Turnover		-0.04	-0.04	-0.04
		(0.02)	(0.02)	(0.02)
Volatility		1.02	1.07	1.07
		(1.27)	(1.28)	(1.28)
Analyst Revision			0.62***	0.62***
			(0.09)	(0.09)
Earnings Surprise			2.60***	2.60***
			(0.51)	(0.51)
Tweeted Day	-0.09**	-0.07	-0.01	-0.13
	(0.04)	(0.04)	(0.04)	(0.20)
Const.	0.03**	0.01	0.01	0.01
	(0.01)	(0.04)	(0.04)	(0.04)
News Sentiment	No	No	Yes	Yes
Obs.	12633155	12633155	12633155	12633155